Z 699
.K45
Copy 3

# AUTOMATION and the LIBRARY of CONGRESS

Fiscal Services

# AUTOMATION AND THE LIBRARY OF CONGRESS

A survey sponsored by
THE COUNCIL ON LIBRARY RESOURCES, INC.

*Submitted by:*

GILBERT W. KING, *Chairman*
HAROLD P. EDMUNDSON
MERRILL M. FLOOD
MANFRED KOCHEN
RICHARD L. LIBBY
DON R. SWANSON
ALEXANDER WYLLY

LIBRARY OF CONGRESS • WASHINGTON : 1963

L.C. Card 64–60015

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402 - Price $2.00

# Contents


| | Page |
|---|---|
| Letter of Transmittal | v |
| Acknowledgments | vii |
| Section: | |
| I. The Automation Survey: Background and Conclusions | 1 |
| Background | 1 |
| Conclusions | 2 |
| Recommendations | 2 |
| Benefits of Automation | 2 |
| II. Assessment of Need | 4 |
| The Research Library Today | 4 |
| Automation's Promise | 6 |
| III. Assessment of Feasibility | 7 |
| General Requirements | 7 |
| Storage and Processing Requirements | 7 |
| Multiple-Access Requirements | 8 |
| Printing Requirements | 8 |
| Feasibility of Storage, Processing, and Input | 9 |
| Feasibility of File Conversion | 9 |
| Microform Storage | 9 |
| Feasibility of Automated Stack Storage and Retrieval | 10 |
| State of Information Science | 11 |
| Conclusion | 11 |
| IV. Projection of a System | 12 |
| Operational Description | 12 |
| Cataloging | 12 |
| Use of Consoles by Library Clients | 13 |
| Section—Continued | |
| IV. Projection of a System—Continued | |
| Operational Description—Continued | |
| Use of Consoles by Library Staff | 15 |
| Methods of Search | 15 |
| Search Tables | 16 |
| Use of Microform Storage | 17 |
| System Characteristics | 17 |
| V. Implementation of a System | 19 |
| Management of the Transition | 19 |
| Auxiliary Research and Development Needs | 19 |
| Trillion-Bit Memory | 19 |
| Consoles | 19 |
| Graphic Output Editing and Composing | 20 |
| User-System Tutorial Interaction | 20 |
| Communication Technology | 20 |
| Textual Data Conversion Methods | 20 |
| VI. Consequences of Automation | 21 |
| Impact on the National Library System | 21 |
| Information Flow | 21 |
| Interlibrary Cooperation | 21 |
| Impact on the Library User | 22 |
| Effect on Availability of Scientific and Technical Information | 22 |
| Organization of the Collections | 23 |
| Measurement and Feedback of Library Effectiveness | 23 |
| Concluding Comment | 23 |


## APPENDIX


| | Page |
|---|---|
| A Cost Analysis of an Automated System for the Library of Congress | 27 |
| Foreword | 29 |
| Section: | |
| I. Background and Summary | 31 |
| Background | 31 |
| Summary of Findings | 32 |
| Conclusion | 33 |
| II. Identification of Major Library Activities | 35 |
| Summary | 35 |
| Activities Considered for Cost Analysis | 35 |
| The Processing Department | 37 |
| The Reference Department | 38 |
| Section—Continued | |
| III. Sketch of Functions to be Performed by an Automated System | 40 |
| Summary | 40 |
| Acquisition and Recording Functions | 40 |
| Cataloging Functions | 41 |
| Staff and Reader File Access | 42 |
| Circulation and Control Functions | 43 |
| Editing Functions | 44 |
| Administrative Functions | 45 |
| Miscellaneous Functions | 46 |
| IV. Statistics for Key Library Operations | 49 |
| Summary | 49 |
| Introduction | 49 |

Section IV—Continued Page
Acquisitions and Internal Processing Activities .... 49
Cataloging Activities .... 51
Catalog Use .... 51
Circulation and Retrieval Activities .... 53

V. Basic Hardware Specifications .... 54
Summary .... 54
Introduction .... 54
The Central Catalog .... 54
The Monograph Processing File .... 56
The Serial File .... 57
The Exchange Record .... 60
The Location Directory and the Not-on-Shelf File .... 60
The Circulation File .... 61
The Catalog Authority File .... 61
The Combined Function Groups .... 62
Consoles .... 66
Hardware Requirements for Editing and Publication Functions .... 69
Hardware for Graphic Storage and Retrieval .... 69
Other Hardware Requirements .... 70
Functions Excluded or Analyzed Separately .... 70
Résumé of Hardware Specifications .... 71

VI. Hardware and Software Costs .... 73
Summary .... 73
Hardware Costs .... 73
Software Costs .... 76
Rental vs. Purchase .... 77

VII. Operating Costs of the Automated System .... 78
Summary .... 78
Personnel Costs .... 78
Recurrent Hardware-Related Costs .... 80

VIII. Costs of Conversion to Digital-Form Storage .... 82
Summary .... 82
The Conversion Process .... 82
Cost of File Conversion .... 83

IX. Cost Projection for the Manual System .... 85
Summary .... 85
Introduction .... 85
Salary Cost Estimates .... 85

## *List of Exhibits*

1. Functional Block Diagram of an Automated Library System .... facing 31
2. Functions of an Automated Library System .... 31
3. Procurement Costs of an Automated System .... 32
4. Organization Chart of the Library of Congress .... 36
5. Summary of Departments with Respect to Cost Analysis .... 37
6. Divisions of the Processing Department .... 37
7. Divisions of the Reference Department .... 39
8. Statistics for Acquisition and Processing Activities: 1949–61 .... 50
9. Serial Record Processing Statistics: 1953–61 .... 50
10. National Union Catalog Statistics: Cards Submitted by Other Libraries: 1949–61 .... 51
11. Estimated Catalog Uses for Reference Services: 1949–61 .... 52
12. Card Division Statistics: 1949–61 .... 53
13. Statistics for Volumes Delivered, Retrieved by Photoduplication, Shelflisted, and Lent: 1949–61 .... 53
14. Central Catalog Specifications: Initial Size .... 55
15. Central Catalog Specifications: Number of Uses Per Year .... 56
16. Central Catalog Specifications: Number of Uses Per Hour and Per Second .... 57
17. Monograph Processing File Specifications .... 58
18. Serial File Specifications: Capacity and Use .... 59
19. Exchange Record Specifications .... 60
20. Location Directory Specifications .... 61
21. Not-on-Shelf File Specifications .... 62
22. Circulation File Specifications .... 63
23. Catalog Authority File: Specifications .... 64
24. Combined Function Groups .... 65
25. Console Keyboard Character Set .... 66
26. Preset Message Formats and Operation Instructions for the Consoles .... 67
27. Number of Consoles Required .... 68
28. Summary of Hardware Specifications and Requirements by Function .... 71
29. Basic Hardware Costs of the Automated System .... 74
30. Sources of Price Estimates for Basic Hardware Items. 76
31. Summary of Basic Hardware and Software Costs .... 77
32. Annual Salary Costs of the Automated System: 1972. 79
33. Description of Hardware-Oriented Groups .... 80
34. Recurring Annual Costs of Basic Hardware Items .... 81
35. Files to be Converted to Digital Form .... 84
36. Costs of Conversion of Existing Files to Digital Form .... 84
37. Salary Costs of the Manual System in 1972: Acquisition and Internal Processing .... 86
38. Salary Costs of the Manual System in 1972: Cataloging and Catalog Maintenance .... 87
39. Salary Costs of the Manual System in 1972: Circulation, Retrieval, and Storage .... 88

DECEMBER 1, 1963.

The Honorable L. QUINCY MUMFORD,
*Librarian of Congress.*

DEAR SIR:

We are pleased to be able to submit to you a report in which the automation of major operations within the Library of Congress is shown to be both desirable and feasible. The main body of the report argues that automation is needed because, with present methods, research libraries are failing to meet developing information needs not only in the areas of science and technology, but also in the arts and humanities. We believe that libraries should take advantage of modern technology in order to offer new, and hitherto unrealizable, services to their users.

The Appendix to the report presents an analysis indicating Library functions that can, and should, be automated and what the costs are likely to be.

Based on this study, we are unanimously agreed that immediate action should be taken to establish a program leading initially to automation in the Library of Congress and ultimately to the incorporation of large research libraries throughout the country into an automated library system.

Respectfully submitted

GILBERT W. KING, *Chairman*
HAROLD P. EDMUNDSON
MERRILL M. FLOOD
MANFRED KOCHEN
RICHARD L. LIBBY
DON R. SWANSON
ALEXANDER WYLLY

# *Acknowledgments*

The survey team acknowledges its appreciation of many formal and informal discussions on the problem assigned to it. In particular, the team appreciates the guidance and support given to it personally by Verner W. Clapp, President of the Council on Library Resources, Inc.

The aid given to the survey team by the staff of the Library of Congress and, in particular, by Henry J. Dubester and Mrs. Barbara Evans Markuson in almost every aspect of this survey has been of enormous value. They imparted to the team an insider's feeling for Library problems, organization, and practices, and they gathered needed data and undertook a number of special investigations for the team. Their comments upon the initial draft of the report did much to improve its final version.

It has been a rewarding personal experience to work with the Library on this project. L. Quincy Mumford, Librarian of Congress, and Rutherford D. Rogers, Deputy Librarian of Congress, enhanced the work of the survey team, not only by giving so much of their own time but by encouraging an atmosphere of cooperation and interest within the Library toward the survey and the team members.

The survey team also wishes to express its appreciation to the members of its Advisory Committee, Herman H. Fussler, Director of the University of Chicago Library; Edward M. Heiliger, Director of Library Services, Florida Atlantic University (formerly Librarian of the Chicago Undergraduate Division Library, University of Illinois); Frank B. Rogers, Librarian of the Charles Dennison Memorial Library, University of Colorado Medical Center (formerly Librarian of the National Library of Medicine); and Frederick H. Wagman, Director of the University of Michigan Library. Their ability to interpret the modern research library milieu to outsiders and their comments and suggestions have been of great value throughout this project.

David F. Loeb of the Itek Corporation has contributed to the editing of this report.

Much of the work of the analysis presented in the Appendix was performed by Herbert T. Spiro and Allan D. Kotin of the Planning Research Corporation, to whom the survey team is greatly indebted.

SECTION I

# *The Automation Survey: Background and Conclusions*

## *Background*

During the past decade the Library of Congress, in common with many other research libraries, has become increasingly aware of pressures and strains in many of its operations. These areas of concern (well known to librarians) include mounting arrearages in cataloging and other processing activities; increasing complexity of manual inventory control files; increasing difficulty in keeping card catalogs reasonably current and accurate; and increasing demands for a wide variety of services for a clientele whose needs for information have grown tremendously since World War II, and many of whom have only recently turned to large research libraries for help. Five years ago an internal committee of the Library of Congress was charged to study potential applications of electronic data processing equipment to Library procedures; subsequently representatives of three computer firms made brief studies of Library operations which intimated that certain areas could benefit substantially by automation.

The Library took the stand that it was not immediately concerned with mechanizing a few operations; it sought a plan to provide a blueprint for its actions during the next 5 or 10 years. To do this, the Librarian, L. Quincy Mumford, asked the Council on Library Resources, Inc., on behalf of the Library of Congress and other large research libraries, for help in conducting a study to consider the practicability and advisability of applying mechanization to the total bibliographical system of a large research library. The approach suggested to the Council was a survey conducted by a team of outside specialists whose efforts would be supported by a staff within the Library to aid in the necessary studies and to provide liaison between librarians and survey team members. Because it was most desirable to reflect the needs of other libraries whose problems may not be entirely identical with those of the Library of Congress, it was suggested that an Advisory Committee of librarians from other libraries be selected to provide the survey team with an expert group whose experience could be drawn upon when needed.

On April 23, 1961, the Council on Library Resources, Inc., announced a $100,000 grant to the Library of Congress for "a survey of the possibilities of automating the organization, storage, and retrieval of information in a large research library . . . not only from the point of view of the functioning of an individual institution but also from that of a research library whose activities are interrelated with those of other research libraries."

To undertake this investigation the Librarian of Congress invited Gilbert W. King to head a survey team of technical specialists. Other members named to the seven-man team were Harold P. Edmundson, Merrill M. Flood, Manfred Kochen, Richard L. Libby, Don R. Swanson, and Alexander Wylly. To the Advisory Committee Dr. Mumford appointed four distinguished librarians: Herman H. Fussler, Edward M. Heiliger, Frank B. Rogers, and Frederick H. Wagman. Verner W. Clapp, President of the Council of Library Resources, Inc., participated in the initial planning of the survey. Henry J. Dubester, Chief of the Library's General Reference and Bibliography Division, was detailed to serve as coordinator of the program, and Mrs. Barbara Evans Markuson was appointed to assist him and the team.

As the Library of Congress itself was to be the principal focus of the study, an intensive orientation period was provided by means of reports, demonstrations, and meetings concerning various Library operations. Arrangements were also made for the team to visit, as a group, the University of California Libraries in Los Angeles and in Berkeley. Individual team members visited many other large libraries both in the United States and abroad. As required, special analyses were made either by team members, by the Library supporting staff, or by contract to an outside firm.

715-690—64——2

One such report is included in the Appendix to this document. (Studies prepared by the Library's Office of the Information Systems Specialist and the preliminary costing study conducted by Planning Research Corporation were developed for internal use only; copies are not available for distribution.)

The grant provided funds to support a total of 3 months of investigation spread over as many months as it was deemed necessary. According to the grant from the Council, the survey was expected to result in a statement "of the feasibility of mechanization of research library activities and of requirements for such mechanization." The report obviously cannot do more than provide guidelines for consideration by the Library administration. The team, in accordance with the stipulations of the grant, considered both current technology and foreseeable developments. This report is not, and should not be considered, the result of a detailed systems design study which it, indeed, recommends as the next step. It is a study of the feasibility of the application of automation to certain functions of the Library of Congress and, as such, reflects the opinions of the survey team members studying this problem.

## *Conclusions*

The survey team reached the following conclusions:

1. Automation can, within the next decade, augment and accelerate the services rendered by large research libraries and can have a profound effect upon their responsiveness to the needs of library users. (See pages 4–6, 12–18, 22–23.)
2. Automation of bibliographic processing, catalog searching, and document retrieval is technically and economically feasible in large research libraries. (See pages 7–10.)
3. The retrieval of the intellectual content of books by automatic methods is not now feasible for large collections, but progress in that direction will be advanced by effective automation of cataloging and indexing functions. (See pages 10–11.)
4. Automation will enhance the adaptability of libraries to changes in the national research environment and will facilitate the development of a national library system. (See pages 21–23.)
5. Automation will reduce the cost-to-performance ratio; however, the Library should aim at the expansion of services rather than the reduction of total operating costs. (See pages 5, 11.)

## *Recommendations*

The survey team recommends that the Librarian of Congress take the following actions:

1. Request $750,000 to be devoted to securing system specifications for the automation of the internal operations of the Library of Congress and the functions it performs for other libraries.[1]
2. Request funds for implementing the system specified by the above design effort immediately upon its completion.
3. Establish a group within the Library to administer the automation project and to assume responsibility for its continuing implementation.

With respect to the second recommendation, it is impossible to arrive at a precise and accurate total cost for the automation project prior to a detailed design effort. The cost study included in the Appendix was made at the direction of the survey team to establish bounds. Based on this study, the survey team estimates in a preliminary way that successful automation of the scope envisioned in this report can be carried out for a sum between $50 and $70 million, an amount approximately equal to the budget of the Library for 3 years.[2]

## *Benefits of Automation*

The immediate objective of automation will be to solve the pressing problems that face research libraries, among which are problems of bibliographic organization and control. In the long run, however, the most

[1] This amount is based on the team's experience with the cost of similar projects and their estimate of the number and types of personnel required to work on various aspects of the study.

[2] The cost study, reported in the Appendix, estimates that the procurement of hardware, software, and necessary file conversion for automation of the central bibliographic operation of the Library of Congress would amount to about $30 million. The survey team accepts this as a reasonable cost for the system described; however, it strongly urges that the Library of Congress go beyond the minimum automation of its operations and work toward a network of libraries which it serves in a central capacity. It should work toward the development of a communications network among research libraries, the development of automated graphic storage systems, and the automation of the card distribution services (excluded in the cost study except for some minor operations). The team believes that the $50 to $70 million estimate is the more realistic to achieve an automated library system whose benefits can be shared by research libraries throughout the United States.

significant effect of automation will be the focusing of the services of the library on the individual user for the optimal satisfaction of his research needs. Ideally an automated system should place the full resources of the library at the immediate disposal of the user. These resources should be more varied in nature and contain a much richer set of statistical and bibliographical data than it is practical to provide in conventional libraries. The services of reference librarians can be greatly extended in those areas where the fallibility and limit of human memory are a barrier to providing information that the user needs. At the same time our relatively scarce resources in terms of expert reference librarians will be better utilized in those areas where human judgment and intellect can provide services beyond the capabilities that can be supplied by a mechanized system.

Further, the benefits of the automation of the Library of Congress may be shared by all libraries that elect to use the Library's services.

Some specific short-range benefits of automation will be to:

1. Accelerate the acquisition and processing of library materials and permit effective file monitoring.
2. Increase the reliability and completeness of response to requests.
3. Decrease the effort of user and librarian in the search for information and simultaneously provide more rapid service.
4. Accelerate the production of bibliographies and other library publications.
5. Provide greater assurance of the integrity of the collections.
6. Permit increased depth of indexing.
7. Improve the efficiency and reduce the cost of serial record control, circulation control, and other record-keeping functions.

The most important longer range benefits will be to:

1. Increase use of libraries in support of nationwide endeavors.
2. Service information requests of greater complexity and with a higher degree of responsiveness than is currently possible.
3. Create information compendia tailored to the user's needs.
4. Enhance interlibrary cooperation and secure for the individual user the benefits of the community of library resources.

The sections that follow amplify the above conclusions and recommendations and describe the phases of implementation, including a transition period.

SECTION II

# *Assessment of Need*

## *The Research Library Today*

One of the notable changes in our culture over the past 2 decades has been the growth in academic and technical activity. The number of people drawn into this activity has increased at least tenfold. The effect has been to make information a critically important national product. By its nature, information is heavily dependent upon itself, that is, upon prior information; thus, improving its use can have tremendous implications in our future academic and technical growth. If automation can bring about such improvement, it will surely be justified by its contribution to the quality of a large and important part of our national effort.

Research librarians face a range of serious problems which stem from the rate at which information is accumulating and from the changes in its production and use in our society. For example, almost every research library now has some arrearages in the processing of newly acquired materials. In some research libraries these arrearages constitute a very serious backlog of partially processed or unprocessed material. The acquisitions procedures of research libraries have become ponderous and slower than is desirable because of the mass of detail that must be managed. Serial publications present increasingly complex problems in ordering, the prompt claiming of missing issues, and the fiscal management and bibliographical identification of titles; yet serial publications are rapidly increasing in number and in importance for research. There is a need for more intensive subject analysis of periodicals and other kinds of research material than is now possible for economic and other reasons. Concomitant with the growth and size of library collections is the growth in size and complexity of the card catalog which is getting more and more difficult to use merely because of its size and dispersion. Reorganization of the card catalog and the book collection are almost impossible under present conditions.

Most libraries have severe problems of space; buildings are relatively inflexible structures and accommodation to the growing collection is difficult. Any attempt to make a useful distinction between widely used material and little used or useless material is very difficult within the framework of present techniques and practices. Present library systems cannot accommodate easily to providing a range of response times geared to the requirements of their users. Requests which should receive responses in a matter of minutes are generally handled by the same methods and techniques as those for which a delay of a day or two is acceptable. Manual techniques make it altogether impractical to provide the very rapid response required in many instances of manipulating and examining catalogs or bibliographies. Moreover, the size of card catalogs has so strongly dictated two or three, rarely four or more, access points to a given work, that rigid cataloging systems (both subject and descriptive) have persisted. Subject headings and subject heading lists cannot be maintained at the level of currency required for effective retrieval. Finally there is too much duplication of effort among the various research libraries that make up our national capability for communicating recorded knowledge. This multiplication of effort pervades the cataloging process, the production of bibliographies, the function of acquisition, and even the contents of the collections.

In science and technology the technical report literature has not found its way, other than by exception, into the general corpus of research library materials and services. Almost all of the subject analysis of the scientific and technical periodical literature is provided by services operated outside research libraries. Certain classes of literature, including reprints, preprints, specialized reports, abstracts, etc., are increasingly collected and circulated by decentralized and subject-specialized information centers. It is appropriate to raise the issue as to what extent specialized and tailor-made information services should be provided by the general research library if it is to avoid the role of a mere repository or a passive adjunct to the academic process.

The resources of research libraries should be used more extensively than they now are in a great many of the present research activities in the arts and sciences, but in order for this to happen new technological aids for the library must be introduced.

There is no simple measure of the usefulness of libraries, but their effect is observable in the quality of our intellectual endeavors. Large research libraries, mainly centered in the academic community, have always played a vital role in the training of succeeding generations of leaders in every field of knowledge and enterprise. Libraries support academic research. They are the adjunct of our expanding culture and provide the connective with other cultures now and in the past. Despite their importance, however, they generally have had extremely limited resources made available to them. The burden of effective exploitation has thus always been, and will always be, on the user himself, but substantive techniques for lightening this burden have fallen far short of what could be done with increased resources. The "information explosion" has resulted in disproportionate increases in the complexity of subject interrelatedness, so that the user is faced with penetrating an almost impossibly complex reference maze that is growing more complex. The need for greater depth of subject control has made it increasingly difficult for librarians to organize their materials for the user's benefit, and the specialized indexing and abstracting services have filled only a part of this need. To some significant degree, a lack of rapport between large research libraries and the new class of information users has developed. For example, numerous specialized documentation centers have been established in industrial organizations. The cost of this, in duplicated acquisitions alone, is very high, and the effect on intellectual activities may be profound.

It is possible to identify many different kinds of library users and user needs. They range from the individual who enters the library for intellectual recreation or self-improvement to the investigator who requires answers to a specific question stemming from his research. More often than not the investigator knows that he is unlikely to find a specific answer, but he goes to the library to obtain clues and suggestions. It is generally assumed by librarians that, within wide bounds, all purposes and interests have equal value, so that the serious research worker often competes for library service on virtually equal terms with all other users. In the future there should be greater focus upon the user who requires library facilities for support of his research work, whatever the field of endeavor. As libraries become more responsive to research needs, they are certain to be used much more extensively, and with greater efficiency, for research purposes. Thus, the total cost of automation must be judged against the anticipated, rather than the present, experience of library use.

For the most part, the library today has completed its task when its materials are stored in an orderly fashion and when an item with probable relevance to the subject of inquiry can be identified and delivered to the user. The library of the future should be one that actively participates by predicting the areas for which information will be needed and in minimizing the time the user must spend in ascertaining pertinence of library materials to his special interests. Several years ago optimism prevailed about what could be done in the area of automatic indexing, abstracting, etc., and much of the research in information retrieval currently in progress was started then. This optimism has been somewhat dissipated because solutions have been slow in coming or have not come at all. It has also been recognized that work to solve the more immediate problems of libraries has been underemphasized. In this report the point is made that both the immediate and future problems have a common source—the explosion in research activity—and have to be considered as a whole. The present systems for the bibliographic organization and display of the resources of large research libraries are not sufficiently adaptable to rapidly changing requirements, and they no longer lead the serious investigator easily and directly to pertinent information. Changes are clearly needed if the incalculably valuable information housed in large research libraries is to be put to use by society.

Publication patterns have been changing, particularly since World War II, with the advent of the research report to the Government. Libraries have been criticized for their inability to assimilate new kinds of material within their traditional processing techniques. Also they have faltered, or depended upon other information services, in the indexing of serial literature. Essentially they have retained the book as the focus of their indexing and classification techniques, but books cannot have the currency that the modern tempo in research requires. Even journals are becoming too sluggish. The latest development is the growth in importance of preprints and mimeographed memoranda. Libraries will have to develop capabilities for processing these new materials. Their

nature does not differ in any substantial way from other library materials, but they can be looked upon as added burdens to the functions of acquisition, cataloging, and reference for which library resources are already minimal. Since the unpublished report literature is not subjected to the same editorial sanctions or critical review processes as are published journal articles, it might be argued that these materials are of lower quality and do not deserve the same attention as other library materials. However, since there is much valuable information in the unpublished report literature that is not available elsewhere, this does suggest that more effort must be invested in identifying that portion which is of suitable quality for preservation. Automation can potentially contribute to the problem area of quality identification, descriptive cataloging, and subject indexing, and to storage and distribution through microform techniques.

Because of the nature of modern technology, the benefits of automation would be realized first, and the cost supported, by the very large libraries. The libraries of the United States, and indeed of other countries as well, hope that leadership for the evaluation and the implementation of automation will come from the Library of Congress.

This brings us to the central question of our inquiry: Can automation be of help in an important way? Mechanization of current functions is not primarily what is meant, since, if this were to be the limit of the effort, many users would obviously be neglected because the intellectual environment has changed, and they find themselves with needs not currently satisfied. Moreover, the real potential of automation would be sacrificed. Automation, even in areas where it is well established, can still be expected to yield much more than it has already. In the library, which is as yet virtually untouched, it is particularly important to avoid merely perpetuating past practices. Future possibilities and alternative ways to realize them must be kept in mind at every step of the way. In assessing the usefulness of automation, what we hope to achieve in the future is as important as what we know can be done now.

## *Automation's Promise*

There is a spectrum of functions which automation can improve, ranging from purely routine processing chores, such as circulation control and the maintenance of serial records, to advanced information retrieval techniques. The routine processing operations of large research libraries are complex and are basic to all effective service to readers, but it is believed by the members of the team that the efficiency of many of these functions can be substantially improved with the assistance of automated procedures and equipment, and therefore these matters are not stressed in this report, although some details are given in the Appendix. Rather, emphasis is placed upon those catalog-searching and bibliographic functions which could greatly assist the user and expand the usefulness of large libraries in present-day research activity. A simple example is the assistance automation could give the user in establishing search trails through the various catalogs, indexes, and files. It is most important to observe that the immediate interests and demands of both the recordkeeping and bibliographic functions should be studied within the framework of a single system. There are many potentially serious pitfalls in the independent automation of the many separate functions, since this might result in creating grave difficulties in the ultimate realization of a totally integrated system. In the design study phase, it is recommended that both the routine processing and the assistance-to-search functions be considered simultaneously. Whether or not these are finally implemented with the same equipment is yet to be determined.

What then can be looked for? What are reasonable future goals toward which it is possible now to take at least some first steps? A general goal that influences all others is to make the library a more active organism. The catalogs of libraries must be transformed into dynamic files of records that assist in their own use. Catalogs, however, only direct the user to documents. We should also envision services that go at least one step further, that is, directly to information or at least to fragments or portions of documents. Depth of subject control is at the basis for this hope and it should be recognized that such depth, although facilitated by automation, is not an inevitable consequence thereof. Through improving the flexibility of bibliographic control and through increasing subject coverage in depth, particularly in many areas of science and technology, the possibility emerges that we can add a new dimension to the library's usefulness and to its value as a national resource. Automation will contribute to this new dimension by making practical and economic certain functions that would be extravagant within the framework of existing methods.

# *Assessment of Feasibility*

## *General Requirements*

In terms of information processing technology, the requirements of the Library of Congress may be categorized as follows:

1. Storage of very large quantities of material in digital form (i.e., in a form machines can "read").
2. High-speed processing to manipulate the stored digital information.
3. Storage of vast quantities of graphic material in a form which can be easily and quickly manipulated (principally photographs of pages of documents).
4. File conversion, e.g., the conversion of information now in card catalogs to machine-readable form.
5. Retrieval of the documents themselves, i.e., printed books, periodicals, etc.
6. Communication and display to the user of the various types of information within the system, both digital and graphic, e.g., catalog entries.
7. Output printing, ranging from publication of the National Union Catalog to "throw-away hard copy" of individual bibliographies, etc.
8. Communications with other libraries.

## *Storage and Processing Requirements*

The distinctive feature of the Library of Congress, from the viewpoint of its impact on automation, is the enormous size of its holdings, totaling almost 42 million pieces (as of June 1962), and requiring approximately 270 miles of shelving and containing the equivalent of perhaps $10^{14}$ bits of information.[1] The present catalog files for this collection, if converted, would require $10^{11}$ bits of memory capacity;[2] a memory capacity of $10^{12}$ bits would probably be required to describe the Library's holdings at the anticipated rate of growth through the next generation. Vastly improved image storage in microphotographic form will also be necessary, for a long while, to complement the descriptive digital store. Here the objective should be to make microform more convenient for the user than a book. These demands for storage are a challenge to the information-processing industry.

On the other hand, the rate at which the stored material is changed is relatively low. Currently, on the order of $10^7$ (at most $10^8$) bits are changed or added per day in the course of Library operations[3] as compared to $10^{10}$ bits which are merely looked through without alteration.[4] Thus the processing re-

[1] A "bit" is the unit in which information is measured in communications and data-processing activities. It is an element in the simplest (two-element) code and is easily represented in a machine by the on-off of a switch, the presence or absence of magnetization, etc. To encode digitally an average English word requires about 30 bits. A few extra bits, as prefixes, are required to accommodate other alphabets, font styles, etc.

The Library collection consists of 42 million pieces of which 13 million are books, bound newspapers, and serials; 18 million are manuscripts, leaving a remainder of material as diverse as maps, microprints, music, and photographs. To determine the total contents of the collection an arbitrary estimate of 100 pages per item and 1,000 words per page was made. This gives a total of $4 \times 10^{12}$ words in the Library. At a minimum of 5 characters per word, this would amount to about $10^{14}$ for full conversion of the *text* of the contents of the Library of Congress.

[2] Exhibit 35 in the Appendix indicates that to convert the 50 words of the average main entry card for each of the estimated 6,600,000 titles in the Library of Congress catalog would require conversion of $2 \times 10^9$ characters. At a minimum of 6 bits per character the digital conversion would require $10^{10}$ bits; additional index files required by the system would yield a figure of at least $10^{11}$.

[3] According to the analysis presented in Section IV of the Appendix about 4 million operations involving record alteration are now performed each year in the Library. (These operations include posting serial receipts, maintaining circulation records, etc.). With 3,000 hours of operation per year, $1.3 \times 10^3$ such operations are performed hourly. If we assume that for each operation an average of $10^3$ bits of information is changed, then $1.3 \times 10^6$ bits of information are changed each hour; in an 8-hour workday this would amount to at least $10^7$ bits of information altered or added.

[4] According to figures used for the cost study, the present catalog use is about 50 per minute for look-up operations. If we assume that one card is selected for each use and assume that 1,500 bits of information is average per card, then about

quirement, though certainly not trivial, will be much less difficult to meet than the requirement for storage and access.

## *Multiple-Access Requirement*

The diversity of users of libraries has already been noted. As the files and contents of traditional libraries are divided into a large number of separate physical units (cards, books, etc.), each user has somewhat independent access and the opportunity for an independent adjustment in his mode of use. The preservation of such flexibility in an automated library, where an enormous amount of information must be concentrated in one or only a few machines, is necessary, but will pose an extremely difficult problem of access.

One approach to the problem is to design the central machines, stores, and processors to meet concurrent demands from a multiplicity of terminals, with appropriate equipment at such terminal points. The terminal sets will have to be simple to use, yet provide a variety of possible interactions with the system. Many of these interactions will be very simple, for example, the request for a specific identifiable book, and for these a simple keyboard will suffice as a console. Other types of interactions will be of a more complex nature and will require an extensive dialogue between the user and the machine system. The console for this purpose will include a display device of some kind (possibly a cathode-ray tube) connected to the digital display or the microform viewer. It is possible that a third type of console intermediate in complexity between the two just described will be required.

A detailed study must be carried out to determine the relative importance, in terms of the workload, of the various console functions to be performed by the user. The system should then be designed to minimize the idle time of all consoles and to avoid the use of expensive or complex consoles for carrying out relatively simple operations. The use of consoles should be avoided altogether, when it can be done without loss of response time and flexibility.

A large fraction of the bibliographic product of the Library of Congress can be produced on an assembly-line basis with printed catalogs and bibliographies as the output. For many requests placed on the system by users, a printed catalog may serve the purpose as well as a console, and the choice as to which facility is provided should then be based largely on economic factors.

Most consoles should have a simple intercommunication system or telephone to provide communication with reference librarians. Reference librarians should, in turn, have access to other librarians in the country by means of a communication network, for example, a teletype.

## *Printing Requirements*

The Library of Congress provides a valuable service to other libraries through the distribution of its printed catalog cards, lists of holdings, and book catalogs. The availability of an automated catalog will permit the computer to be used to compile special bibliographies and to produce, as required, specialized catalogs of various kinds.

Where depth of subject control is desirable, these catalogs or bibliographies should exploit to the extent practical the products of subject-specialized indexing and abstracting services. Printed and bound supplements to *The National Union Catalog* should be issued bimonthly and cumulated at least quarterly, yearly, and on a 5-year basis. Production quantity should be between 1,000 and 5,000 in order to provide multiple-copy distribution to 60 or so major research libraries and single-copy distribution to other research libraries and public libraries.

Under the present system, subject headings are continually being added or revised, but the subject heading list incorporating these changes is printed only every few years, with supplements issued between editions. This list should be revised and reprinted yearly; distribution in the hundreds is required.

The printing and production of specialized bibliographies will be another major requirement of the system. Many of these bibliographies will be produced in editions of less than 100 copies and may vary in length from 1 to 50 pages. It is roughly estimated that possibly up to 1,000 per day would have to be produced. Further detailed study is necessary in order to determine this volume.

Part of this demand for special bibliographies will be met directly at the user console, where a brief bibliography may be produced in hard copy as an immediate response to a request. The bulk of the bibli-

---

$8 \times 10^4$ bits of information are selected each minute, or about $4 \times 10^7$ operations per day. With the reasonable assumption that to select the desired card at least 100 bits of information are scanned on 10 cards, then $10^{10}$ is a reasonable approximation of the present daily look-up operation.

ographies produced, however, will be on a time scale of a half day or so, rather than on an immediate demand basis. In the Appendix to this report data are given, on the basis of which volume estimates of the required printing load may be made. (See page 69.)

Legibility to permit the rapid scanning of the printed product is an important requirement that implies a capability for printing in a variety of type fonts and sizes. Output printers for computer systems have adequate speed but poor readability; for the most part, they print in a single style and only in uppercase letters. A variety of type styles, weights, and sizes are needed to meet the standards of Library publication.

A remote printing capability may also be needed to provide rapid access to holdings for libraries using the Library of Congress as a repository; however, a careful study must be made of response times actually required, since quite possibly they can be met with centralized printing and mail distribution.

The requirement for high-speed, high-quality, multifont printers has been recognized by several designers of other systems both in Government and commerce. It is believed that equipment meeting the needs of the proposed system for the Library will be available in the near future.

## *Feasibility of Storage, Processing, and Input*

Within recent years technology has addressed itself to the foregoing problems; the technical feasibility of meeting all the requirements outlined can be demonstrated in development laboratories.

Specifically, it is now feasible to meet storage and access requirements for catalogs and indexes suitable for an automated Library of Congress system (from $10^{11}$ to $10^{12}$ bits). Equipment with rapid access to over a billion bits is already commercially available, and it is a reasonable assumption that the demands of a library system could be met in the near future at a reasonable cost.

The data-processing requirements (on the order of $10^9$ bits daily) are already being met in somewhat similar systems that must provide "on demand" services. The terminal sets, for both input and output functions, are technically feasible, but they need more development work to be adapted to the functions of the Library of Congress system.

Communications media for the transmission of digital and graphic material have been developed recently at the technical level to be practical for the system visualized, and a variety of transducers for conveying information between terminal sets or data processors and communications networks are now available.

## *Feasibility of File Conversion*

The process of converting information from textual form to machine-readable form, which is necessary for both the transition to an automated system and its continual updating, is itself a subject of considerable importance. This conversion could be accomplished with traditional methods of keypunching or typing with simultaneous punching on paper tape.

The conversion of catalog cards is an enormous task; the National Union Catalog, for example, contains about 15 million cards with an average of approximately 50 words per card. Running text can be keypunched and verified at roughly one cent per word. At this rate the complete conversion of the National Union Catalog would cost about $7.5 million. However, the heavy mixture of numerals and the necessity for a certain amount of formatting, error correction, and quality control suggest that the one-cent-per-word figure may be somewhat low.

The automatic conversion of stenotyping is being actively pursued with the aim of reducing the cost of file conversion. Automatic print-reading devices are under development and may at some future date provide a partial solution to the problem.

The technique of file conversion requires further research and development. This activity will undoubtedly result in a method of converting the Library's catalogs into machine-readable form in the near future at an acceptable cost.

## *Microform Storage*

The microform storage of page images and graphic records has been addressed traditionally to preservation of materials, saving of space, and facilitating the duplication and distribution of library materials. The miniaturized record is intended for human consumption and of itself has nothing to do with the machine processing of information.

The problems of cataloging, indexing, and classifying, as well as the coding and recording of such information, can be considered independently of the question of microstorage. To be sure, there are systems of microform storage which combine both a digi-

tal code and a graphic image in a single record, but these systems represent particular approaches which must be evaluated on a speed and cost basis when a detailed systems design study is undertaken. It should be noted that text words in stored micro-images are not accessible to machine processing any more so than are words on a printed page.

The issue of microform storage and its applicability within large research libraries depends strongly on the question of economics. The expense of storing a bound full-size book must be compared with the expense of recording it in microform plus producing enlarged hard copy as necessary. The speed of access in the two cases must be compared with the requirement for access.

Requirements for access time cannot be explicity formulated at this stage, but certain observations can be made. Immediately following a catalog search, or the production of a brief special bibliography, the requester may wish to examine certain portions of books before requesting the books themselves. Portions might include title page, table of contents, index, or perhaps a few selected pages of text which would permit judging the level of technical detail or other characteristics of the work.

Ideally, this graphic material should be delivered for perusual within seconds. If a batch of contiguous material is requested, such as successive pages of a report, then a more stringent requirement exists for sequencing back and forth from a page in about 1/5 of a second.[5] To make a few essential pages from each work available, and with only partial selection of books, journals, newspapers, and reports within the Library of Congress, it will still be necessary to record upwards of 5 million images, and perhaps eventually several times this amount; any given page-image must be accessible within several seconds.

Microstorage for certain types of archival material or complete books will have access-time requirements that may vary from minutes up to a day or so. The volume requirement here is difficult to estimate, but will depend primarily on the economic balance between the cost of hard-copy storage and the cost of miniaturization and subsequent enlargement. This basic form of storage has long been used in libraries, but with primitive mechanization and very slow access.

[5] To the reader a delay of more than 1/5 of a second is sufficiently annoying to reduce the appeal, and hence the use, of the system.

In the past decade great strides have been made so that the technology is certain to be of greater importance to libraries than it has been. Automatic stores holding millions of images are available. Their drawbacks are low speed and high cost. Images can be located and reproduced from these stores in seconds, but even this may be too slow for compatibility with other aspects of automation and the mode of use envisioned. The feasibility of devices in the required speed range has already been proved in research and development laboratories and with technology that does not appear to be inherently expensive.

## *Feasibility of Automated Stack Storage and Retrieval*

A significant feature that poses difficulty in most large research libraries is the inflexibility of the "marriage" between the bibliographic catalogs and the physical arrangement of library holdings. There should be a complete independence of the physical location of items from their descriptive mapping in the catalogs and files. Such independence will encourage the physical storage of material on the basis of probability of use, where this is practical, or in special compact storage, browsing, or reference collections. Automation can facilitate such independence.

There is the possibility that the form in which knowledge is recorded, stored, and distributed will change appreciably within the next decade. Progress in reproduction techniques, particularly from microfilm, has been so rapid that the circulation of most documents in their printed form may become unnecessary. There is also likely to be a trend toward publication "on demand" from text prepared in machine-readable form. With the possibility of such radical changes to take into account, there is some question as to whether or not it would be worth automating the physical retrieval of books from the collection in its present form.

Present methods of physically retrieving library holdings certainly leave much to be desired and will seem worse as automation increases the speed of bibliographic access. The problem can be simply stated. The time required to extract a book from the stacks and deliver it to a central point should be substantially reduced. The advanced state of technology in the area of mechanical manipulation is an inducement for change. Nevertheless, the investment that would be required, in terms of both money for equipment and

effort for the transitional description, appears too great to expend except in new or renovated libraries.

A hidden danger is that such an investment might tend to deter the adoption of the far more significant changes in storage and publication technology mentioned above. Thus, while the desirability for improved physical handling cannot be denied, the exact requirement will depend on the consequences of changes in information handling, which must be given priority. The door, however, should be left open to clever invention. No approach yet proposed appears to have sufficient merit to justify implementation, but new ideas could still change the picture.

## *State of Information Science*

Intimately connected with technical feasibility is the understanding of how information should be organized in the files and processed to answer the demands on the system. The nature of the data is quite different from that which the computing industry encountered in business and scientific work and learned to handle so well. Libraries consist essentially of collections of words in text. These words are combined and have significance to humans that, in our present state of knowledge, cannot be expressed in terms of machine rules. Thus they cannot be normalized or put in uniform format without their character being changed and information lost.

Present automation techniques do not deal adequately with raw text. At the outset only catalogs, inventory files, and indexes should be considered for automation. Even here, to use these in a sophisticated manner, it may be necessary (and there should be no reluctance) to insert manual processing and human decisions when they are essential or too expensive to replace.

The economic and technical feasibility of automatically retrieving information directly from the text of documents has not yet been established and is an extremely complex subject. For this reason, there is likely to be greater emphasis for some years on the retrieval of bibliographic information. The initial objective in the proposed Library system is to use automation to discover which documents are related to the user's information needs. Even during the initial stages, however, automation should allow much richer cross-referencing and descriptive labeling.

A most desirable goal, and a feasible one, is a system that could accept the user's experience as feedback and thus evolve toward greater capability. This could be achieved in several ways. Users' annotations on subject classification or assignment of subject headings could be accepted as input at the console, subjected to further review and editing by a librarian and then incorporated, as desirable, in the system. Users could also comment on the similarity or relatedness of specific papers, reports, or books which they utilize. The idea of accepting user experience as a feedback to the system can be looked upon as amplifying or supplementing citations made by the author. Author citations have traditionally served as valuable reference tools.

A fairly wide range of novel techniques of this kind deserves further study and exploration in order to insure that future automated libraries will be highly flexible in adapting themselves to user requirements.

## *Conclusion*

The automation of the bibliographical control of the Library of Congress is not only technically feasible but could provide a system more responsive to users and more adaptive to changes in user needs, Library operating conditions, publication practices, the national research library system, and information technology.

For some time, however, costs will remain a problem for all the data-processing and microstorage equipment, but they can be predicted to follow a post-developmental curve downward. A further pull downward could be exerted if demands from libraries and similar systems were consolidated to allow a uniform manufacturing attack. A cost study based on actual and estimated prices shows that by 1972 the cumulative costs for automation of the central bibliographic control functions of the Library of Congress system would be comparable to costs for continuing the current manual system.[6]

[6] Reported in Appendix I. Sections VI, VII, and IX of the study are most pertinent to cost problems. Note that this study shows a break-even point by 1972 for those operations included in the $30 million cost estimate. The team recommends that $50 to $70 million be spent to provide, in addition, automation of operations not now performed by the Library. For these operations no comparable manual system exists, nor was one projected for the purpose of cost comparisons.

# *Projection of a System*

In order to assess feasibility and to estimate costs, a partially automated library is described, completely implementable by 1972 and capable of working in parallel with the existing system on a transitional basis by 1969. Many systems may be hypothesized; no claim is made that the one presented is ideal, since designing an appropriate system will be part of the task of implementing the recommendations of the survey team. It is, however, a reasonable system in terms of short- and long-range goals, and is used solely to illustrate the possibilities of automation concretely in terms of present Library of Congress operations.

The survey team identified the following as functions that could be improved by automation, and the proposed system was designed to facilitate these as a minimum.

1. Accounting and management control.
2. Record management for procurement and processing activities, including serials.
3. Development of cataloging aids (authority files, etc.).
4. Cataloging and catalog maintenance.
5. Searching of catalogs and files.
6. Circulation control.
7. Presentation of information to librarians and users.
8. Publication of cards, catalogs, bibliographies, and indexes.
9. Communications with other libraries and agencies.

## *Operational Description*

The projected system consists of a variety of files, heterogeneous in nature, but susceptible to examination with the same equipment and mode of search. A set of input-output consoles allows both user and librarian to have essentially instantaneous access to any file.

The computer program, a part of which may be thought of as linking the consoles and stored files, provides a method of search for each question, adapting the search sequence to the user's needs. Thus, the system can be regarded as both a learning machine and a teaching machine.

Notwithstanding the emphasis given here to the console mode of operation, it should be understood that there are many processing requirements in the system which will not depend upon use of a console. Many, if not most, of the bibliographies produced in multiple copy will be associated with relatively long response times, such as a day or so, or else will be prepared in accordance with standing requirements for which there may well be a weekly cycle. Even longer cycles pertain to the preparation of large catalogs and possibly to large bibliographies. Accordingly, many computer programs for searching the automated catalog, preparing an initial bibliography, editing such bibliography, and finally formatting to produce a printed product must be written.

Systems such as those in operation at the National Library of Medicine and at the Office of Scientific and Technical Information of the National Aeronautics and Space Administration should be studied thoroughly during the systems design phase and the lessons learned therefrom appropriately taken into account. None of the services or products of these centers should be duplicated within the Library of Congress, but operations with similar purposes should be established in areas not covered by such specialized services.

A major task of the systems design effort it to insure that the technology recommmended is suitable for both the console mode of operation and the required assembly-line method of production.

Cataloging. The function of matching an item of information in hand against the information in files and catalogs is one that is present in virtually every library operation. This search and match process is used to determine whether a recommended item is in the Library's collections, whether a received item is the one that was ordered, whether an author entry has been previously established, whether the work is

present in another edition, etc. Undoubtedly cataloging is of prime importance in generating the tool to support this search function. The detailed manner in which the proposed system converges on the cataloging process is shown in Exhibit 1 of the Appendix. The system permits the cataloger to have rapid access to any portion of the authority file, the central library catalog, and other files that he might not otherwise have approached. Searching can thus be reduced to a minimum, freeing the cataloger for the more intellectual aspects of descriptive and subject analysis.

The expansion of subject analysis in greater depth will become possible. This is impractical in existing libraries because of the bulkiness of card catalogs and the inconvenience to humans working with very large files. With an expanded file the necessity for the user to learn the librarian's language may be reduced. Another advantage will be that special catalogs and bibliographies can be produced on a more sophisticated basis, without the need for special procedures. All the responses which the card catalog can now give will be retained, but these do not approach the complexity or variety of questions a research investigator is capable of posing. An automated card catalog can meet more demands, even guiding users in their line of inquiry (as explained in the following paragraphs on the use of consoles).

It should not be supposed, however, that depth in subject analysis is a necessary and inevitable consequence of automation, nor should it be assumed that such depth will be uniform throughout the holdings of the Library of Congress. It should be noted that considerable depth of subject control already exists in many of the specialized indexing and abstracting services.

Even in a future system, however, it will not be economically sensible to attempt to control every page in every book, document, and journal article in the Library of Congress to the same depth as is currently done by these services. Automation will make less costly the expansion of subject control to areas not presently so covered, but, in the last analysis, the decision as to which areas, and how extensively, will have to be made on an economic basis. Furthermore, depth in control should in no case duplicate any of the functions performed outside of the Library of Congress by the more specialized services. Rather, the system should be designed to accommodate the output of these specialized services and to provide similar services in areas not now covered.

Use of Consoles by Library Clients. In the fully operational system, the query console is intended for the individual user. Because some experience will be required for effective manipulation of the console, it would be desirable at the outset to have a reference librarian perform the console manipulation. Careful consideration during the design study should be given to the workload corresponding to the various console operations. With such data, design criteria for several types of consoles, from simple numerical keyboards to intricate display and input devices, can be developed. The following functional description applies primarily to the more sophisticated consoles. User access to the console will provide a catalog-browsing capability which will be enhanced by the ability of the system to provide hard copy of any item displayed on the console screen. Thus, in viewing a given sequence of bibliographic entries, the user will be able to retain copies of those entries which interest him for further consideration or for record purposes.

A capability for relatively elaborate dialogues between the user and the catalog is an important requirement for the more complex consoles. The query-input keys should essentially be of two types, "process keys" that initiate entire computer subroutines, and "data keys" that permit the entry of alphabetical and numerical data. The alphanumeric keyboard should include a local display to permit the immediate checking of keyed information before it is finally entered into the computer system. The displays themselves must be alphanumeric, but a line drawing capability is not required. Particular care must go into the design of the process keys and into the computer programs that they implement, because in this area lie the means to the simplicity and flexibility which are so necessary to the user-console approach. Information displayed in response to commands initiated at the console should provide statistical and bibliographical data of maximum use to the requester in formulating the next stage of his request.

If the query is too broad or vague, the system should not deluge the questioner with all the references available but should indicate to him first the fact that he has asked too broad a question. For example, the reader might ask for an edition of Shakespeare's *Hamlet*. The first response from the system might be that the Library has 374 editions of this work and that a printed bibliography is available. A display on the console screen indicates to the user ways in which he could narrow his query, e.g., by specifying language,

date, type of edition: translation, illustrated, abridged, student edition, etc. The user can then "fill in the blanks" and indicate that he is, for example, interested only in recent English-language student editions. The system would then indicate the number of editions which filled this criteria and, depending on the number and the user's needs, he could further refine his query, scan the catalog entries for the relevant editions on the console screen, or request a printed listing of the entries.

If the reader decides to scan the catalog entries, the equivalent of a catalog card or a set of several cards would be displayed on the console screen. The operator could then press the "Sequence" key and immediately the following "cards" would be presented. This could be repeated until the end of the sequence for the given author's name.

If the operator at any point wishes to refer to a previous card, he may do so by pressing the "Back Sequence" key. Each console will have available to it a buffer file (for example a magnetic disk) that records the information received from the central catalog store. A standard type of buffer might hold the contents of 100 cards (roughly 200,000 bits). Thus, the operator, by a query or succession of queries, will be able to build up a private "card" file, then readily refer to the material he has collected by pressing the "Sequence" and "Back Sequence" keys.

This mode of operation is intended as an example only and other possibilities can be envisioned. A user may choose to specify that only a portion of the information on each catalog card, such as author and title, be listed in order to permit viewing many more at the same time and minimize the back and forth sequencing operation. In effect then, he is presented with a trial bibliographic listing; he may then call for additional information, such as may be present on a complete catalog card, or he may require a display (probably from a microform image) of selected pages of the work itself. The buffer is also suggested as an example only, since alternative ways of providing the required capability may be conceived. The buffer may possibly be an allocated portion of the main memory, which may or may not be time-shared among several consoles, depending upon workload factors.

After the perusal of this visual material, the user may decide to record the information, i.e., copy the "catalog cards." The "Hard-Copy" key may be pressed, with the result that a printout of the information is automatically produced at the console on a piece of paper for the user.

The user may decide to obtain the item indicated on the display from the collection. Several alternatives are possible. In the transition period, he could obtain the hard-copy printout, take it to the call desk, and obtain the item by the present manual method. In a more automated system he could press the "Call" key. The item number or numbers that he designated on the display would automatically go through the system and result in the items being physically delivered to him at or near the console.

So far we have assumed that the requester's requirement was for a specific identifiable book or report and that he knew the author's name. The system must also accept various other kinds of descriptive tags, such as date, language, publisher, etc., in order to present to the user a bibliography of all items which meet a stated request, even though such request may contain only a part of the information normally required in a conventional system.

Another kind of request is one in which the user needs subject access to the library holdings. In this situation he would key in a subject name and press the "Subject File" key. Here we have a more complicated matter, for the user may specify terms which differ from those used by the cataloger. Thus, the first thing the system must do is transform the user's terms, by a process similar to that now used for language translation, into terms meaningful to the system. For example, if "insurgence" were keyed in as a subject term, it might be translated to one or several headings, for example, "insurrection." This term would be displayed to the user for his information and would also augment his original word as the subject of search. In many instances there will be ambiguities. A research worker may be expert in the terminology of his own field without being aware that some of these terms have quite different meanings in other subject fields. If, for example, the user keys in "bonds," the display might show:

"adhesives"
"bonding agencies"
"chemical valence"
"securities"

The user would then select one or more by pressing appropriate keys, which then cause the subject file to be searched automatically. Again, the first display would probably be the number of items under this heading and a suggestion for further restriction of the question.

At first the display will be equivalent to catalog cards or to a bibliographic listing, as noted earlier. As addi-

tional files are assembled and integrated into the system, the catalog will be made a far more flexible tool. In time the system will be capable of suggesting that a given subject is analyzed in a different or deeper way in the catalog of another library. It will also be able to indicate that the subject is well treated in certain encyclopedias, and, as a convenience, it should allow the user to call for the pertinent article from either a digital or microform file and display it for him. It is essential that this system be compatible with the trend to convert documents to film by microphotography. The first use of microform images to extend the library catalog will be in displaying title pages and tables of contents.

Use of Consoles by Library Staff. The console for use by the library staff, in the acquisitions and cataloging operations, may or may not be identical to one of the several types of user consoles, but in any event the requirements for its use are somewhat different. For reference assistance to the user, it is assumed here that the librarians will have access to a user console. To illustrate the use of the staff console, consider the process of descriptive cataloging.

The basic input to the system is the title page data typed in by the cataloger. Edit programs for the processor would be written to execute most of the normalization of the data on the title page to the form of a Library of Congress card. These operations consist of properly arranging information, selecting font styles, referring to the authority file, the National Union Catalog, etc. An accession or item number would be automatically generated and displayed. A preliminary "card" would very rapidly be written by the program and displayed at the console, together with unresolved questions. The cataloger would then edit this material at the console until a perfect Library of Congress "card" would be displayed, at which point, by a control, he would send it to temporary (quick writing) storage. Here it would be used in another program, and material would be developed from its contents for assignment to all appropriate files—author, subject, etc. No typing or preparation of paper documents or cards would be necessary.

Librarians in other libraries would have consoles similar to those at the Library of Congress and would be able to use them for monitoring the status of processing. This would be necessary because, with high-speed capability, the Library of Congress would be able to do the cataloging for almost all new items. For items not in the Library of Congress but held by other libraries, cataloging information could be transmitted for inclusion in the Library of Congress catalog. If this information were fragmentary, it could be fully elaborated by the automated cataloging system at the Library of Congress for retransmission to the initiating library. Furthermore, other libraries in the system could then obtain the complete data, possibly via display consoles if rapid response were needed, in order to prepare catalog cards for their own files. Librarians at the outlying libraries, however, would not have the capability of changing any of the basic Library of Congress files. Further study of remote console use should be carried out during the design of the system since it can be foreseen that a good part of the requirement of communication with outlying libraries can be met by a communications system (e.g., teletype, dataphone, etc.) and with response times much longer than are implied by display consoles.

Methods of Search. So far the system has been described as consisting of two basic elements: means of storing catalog and index information, and means of display, with rapid access, to the user. There is a third element, the implementation of a search procedure. These search procedures may be very simple, but the capability for very complex searches should exist. The system must therefore be designed to accommodate, without radical changes in equipment or undue additional expense, the advent of new search techniques developed either by experience with the system itself or as a result of research in the field of library and information science.

In order to be usable, material within a library must be indexed, classified, and cataloged, by all methods, both subject and descriptive, that are commonly used to gain access to bibliographic materials. This analysis is done either extrinsically by assigning labels, e.g., subject headings, or intrinsically by the direct use of words occurring within the material itself. Under present cataloging practices, titles and authors' names represent the limit of intrinsic indexing. It has to be recognized that to improve retrieval of information more extensive intrinsic indexing is mandatory. In fact, it is desirable for certain types of material that portions of the text itself be made accessible. A minimal requirement for all material in the library is for intrinsic indexing to include at least the present descriptive information, such as title, authors, publisher, imprint date, etc.

With regard to extrinsic indexing, such indexes as shelf number and accession number can be generated automatically. Subject analysis at present requires human effort, and is a processing bottleneck. Never-

theless it clearly must be expanded and improved. It is obvious that present subject analysis by the Library of Congress makes no pretense at coverage in depth. The average number of subject headings assigned to a book is between one and two, even though a dozen subjects may be dealt with. It has been mentioned that any decision to achieve subject control in depth will be limited by economic factors, and that maximal advantage should be taken of output already provided in specialized subject areas by indexing and abstracting services. Certain initial and limited steps should be further studied, for example, recording the tables of contents of books and journals in machine-readable form. An automatic indexing procedure then could permit a reasonably accurate assignment of subject descriptors. Some experimental evidence exists to indicate that this technique, applied to highly technical material, is as successful as human subject analysis in depth of similar material.

A fundamental principle of library systems analysis is that the indexes to the intellectual content of a document be divorced conceptually from the physical storage of that document. As the future automated library is conceived here, index material should be recorded in a memory specifically designed for digital storage. Independence of digital index data from graphic records generally provides a flexibility not otherwise possible, and certainly not practicable at present in the Library of Congress. For example, under the existing system, if a change of classification is desirable, not only the book labels must be changed and all changes reflected in the card catalogs, but the books themselves must be moved. By giving the physical documents merely a permanent accession number and arranging the library by manipulating records within the machine, changes can be made at low cost and high speed. In this context, "changes" means desirable regroupings of index material with growth of the library, with changes in human knowledge and in the analysis of it, and with the advances in information retrieval.

Another important feature of the system, as currently conceived, is that, by proper construction of the files, different schemes of subject analysis, including those of specialized libraries, can be used in the system simultaneously. Then the user of the system will have the choice and be advised of different analyses of the material in which he is interested. A large variety of indexes must be made available; the cost and space requirements of storage in a machine must be studied and compared with what would be required to house and manipulate equivalent card files or book catalogs.

The design objective in automation should center on responsiveness to the user and not on automation as an end in itself. The basic user requirements involve searching for desired material in response to a specific need, and this implies that the files must not be dead lists but must provide their own references while still accommodating decision and choice of path by the user. The files will be dynamic in the sense that the search is automatically transferred from one reference to another until a satisfactory compilation of selective material is achieved. Further study must be conducted on efficient methods for storing and searching for information. This must be done as part of the design study in close coordination with the planning of equipment, since the particular methods of search to be implemented may depend strongly on the equipment itself. The following description of a series of tables in which a search process is referred from one to the next in a serial fashion is intended as an example only.

Search Tables. It is simpler to begin with the final stage of the search, which is the location of the required item. This will be accomplished by a table relating two indexes, the item number (permanently assigned to an item and physically attached thereto), and its current location number. The latter may reflect that the item is checked out or at the bindery. Thus upon looking up an item by its number, the first step will be a reference (e.g., in the not-on-shelf file). The result of this look-up is an automatic reference to the next subsidiary table to be searched—various process files, circulation files, etc.

The identification of an item leading to its location and use can also be accomplished through table look-up. In support of a central file of numbered items described in essentially the form consonant with accepted library practice (but not constrained thereby), there will be tables which relate authors, titles, subjects, and other attributes to item numbers. These tables will be approached (automatically) sequentially to effect identification of single items or groups of items by author, subject, or other discriminating criteria (year of publication, language, etc.). When the author's exact name or the exact form of the subject term is lacking, the system would provide automatic reference to other tables or files which will make the needed names or terms available. In addition there will be

tutorial tables describing to the user facts about the system and suggestions for directing his search.

A feature which must be emphasized in characterizing all of these files is their formal nature as tables giving logical steps to be taken next. Specifically, each entry has an index by which it is searched, an outcome which may or may not be presented on a console, and a reference to the next table which should be referred to, either automatically or at the choice of the user of the console.

Use of Microform Storage. It has been pointed out earlier that the traditional purpose of microform storage has been to save space and permit easier duplication and distribution of library materials. The graphic record produced is for human consumption, not for machine processing; thus provision must be made for rapid mechanical handling and access to these graphic records for presentation to humans.

Depending upon the particular equipment used to implement such access, there may be considerable flexibility in the organization of the microform file in order to accommodate new accessions without a burdensome interfiling process. Items may be entered into the store merely in some convenient order (e.g., chronologically by accession) and assigned a location number, which will be placed in the automated catalog. Thus, after the system finds a desired item in the catalog, the user will have it displayed immediately on the console by specifying its location number. The capability to retrieve microform images of title pages, tables of contents, and indexes, as well as sample pages taken from the work itself, will be exceedingly useful in supplementing catalog searching as part of the browsing process. It is essential therefore that, to meet this requirement, the system respond within a few seconds in sending the images called for. Successive pages of a single work should be presented with a still faster response time.

The nature of the material to be stored in microform covers a wide spectrum, and the limit is one of economics. It may be practical to put catalog cards in this form for some purposes. However, even at the start, it will be necessary to have this information in digital form as well, for manipulation by computer programs. Other reference materials, such as encyclopedias, bibliographies, newspapers, journals, book title pages, chapter headings, abstracts, and indexes, will be stored in microform as the system develops. Ultimately the complete text of many books will be stored in microform, although the balance between this and conventional hard-copy techniques will be governed by cost factors. A great potential of microform storage is that it may permit methods of publication whereby an individual could assemble a tailormade book on any desired topic from a variety of references. The assembly feature will be part of the retrieval program, and editing could be carried out at the console.

The automated library system as projected will provide for only minimal use of microform at the outset but will be designed to accommodate substantial growth in this phase of its operation, as needs, funds, and technology permit.

## *System Characteristics*

The paragraphs above contain some hints of radical changes in library functions. It is therefore worth reemphasizing that the proposed system has as its primary aim the improvement of certain present library functions, specifically those functions listed on page 12. The only gain to be realized initially will be in processing rates and the degree to which holdings are made accessible. As a byproduct, however, the system should generate a complete record of its transactions, thus allowing an evaluation of its performance to an extent that is presently impossible.

The system as proposed will be much more functionally flexible than the present manual system. Basic alterations in the present system are too cumbersome to consider, and even experimentation requires too much effort. With the streamlining of clerical operations and the high speed of response, it will become practical to experiment with parallel analyses and descriptions of the same material. In time the user should be able to search for needed information in catalogs organized from several different points of view, that is, he should have many information systems at his disposal rather than a single system.

There must be compatibility between the Library of Congress and other research libraries, particularly with respect to the kinds of equipment developed. There is some urgency with respect to this compatibility since some efforts at automating individual libraries are already underway. The effort of establishing compatibility, or at least appropriate interfaces, is probably small compared to expected benefits. In regard to the "software," that is, indexing, classification, etc., there is no need to constrain individual libraries. A specialized medical library cannot be expected to organize

its holdings in the same way as the medical collection of a large general library. In an automated system, it should be possible to provide necessary transformations from one system to another. There should be no reason for destroying the user's sense of the history or peculiar qualities of any collection. Unique indexes should not be lost in an overall national indexing system but should be made available to any user of the automatic system.

The projected automated system is designed to permit an orderly transition to more advanced modes of response to the user's needs. As the quality of indexing increases so will the user's resourcefulness in searching. At the outset, then, machines can hardly be expected to give much more than clerical assistance. In time their power will come to be applied more widely to assist the cataloging, indexing, and assembling of information. As machine processes are improved for these functions they will also become increasingly adaptive to the user's needs.

SECTION V

# *Implementation of a System*

## *Management of the Transition*

The first step to be taken in the automation program is the establishment within the Library of Congress of a management group to insure effective control, coordination, monitoring, and evaluation of activities leading to the successful procurement of a working system. This group will procure a detailed system design that will entail flow charting of essential processes and the determination of machine programming requirements, manpower requirements, machine and machine-hour requirements, throughput volumes, and throughput rates. These design activities should result in a set of performance specifications for the system as a whole and for system components as necessary. The development, installation, and testing of the working system will be based on these specifications.

The transition to the new system admits a variety of alternatives, whose choice will be conditioned by consideration of costs of converting all or part of the present catalogs and related files into machine-readable form, by factors of availability of requisite funds and space, and also by certain development requirements that may make parts of the system available at different times. It is envisioned that the new system, or parts of it, will be operated in parallel with the old, since it would be impossible to make the transition in a short period of time.

## *Auxiliary Research and Development Needs*

The implementation of a system at a cost that is not prohibitive will require that emphasis be placed on the following areas of research and development as early as possible. These areas essentially require engineering development rather than basic technological advances insofar as equipment is concerned.

Trillion-Bit Memory. Storage in, and access to, an automated catalog containing on the order of $10^{11}$ to $10^{12}$ bits is a formidable engineering problem, although equipment can be demonstrated in the laboratory which apparently indicates a promising direction for such a development. There is a requirement for further engineering development, but probably no major basic research or new development programs are needed.

Consoles. Display consoles suitable for functions similar to those envisioned here are available now, but they are by no means entirely satisfactory. Their cost is high for library applications, particularly since perhaps thousands of consoles will be needed for the national library system. New ideas, not just engineering or production changes, are required to overcome both cost and performance obstacles in the same development.

Visual display consoles appear to be desirable, but whether they are to be of the cathode-ray-tube type is not as important as whether, at least, the equivalent of a library card (roughly 1,000 characters) can be displayed, and changed, on the order of one second. Similarly, the display method must allow some form of selection by a "marker" from a multiple-choice array to keep user keying to a minimum.

Temporary storage, either at the console or the central computer, of at least 100 library card equivalents should be provided, with the user able to call for a sequential display of the stored contents. A keyboard supplemented with marker and process keys would be required.

The "process" keys would be used to institute queries to specific portions of the automated catalog and to respond to the next steps suggested by the automated system. At least a 120-character or symbol set would be desirable, in contrast with the 64 characters typically available now. This means that effort will have to be placed on new or improved and certainly more economical means of character generation.

There should also be some means, whether at the console or central processor, of automatic format control and a means of automatic transliteration of words

in digital storage to the Roman alphabet, regardless of the source language alphabet. Lastly, economical methods of providing "scratch pad" copies of displayed information to the console user must be sought.

The foregoing requirements refer to the most complex type of consoles. Again it should be emphasized that several forms of consoles should be developed, depending upon the workload associated with various functions required.

Graphic Output Editing and Composing. Automation of the Library of Congress and the National Union Catalog bibliographic information system holds promise for providing catalog production service on a national basis. Machine methods of editing, spelling correction, type selection, page formatting, and line justification have all been demonstrated. Stress must be placed on software techniques for editing and page composition. Computer graphic composer equipments must be made capable of providing high quality, reproducible copy at rates of several hundred characters per second with a large repertory of symbols, characters, and type sizes.

User-System Tutorial Interaction. Both console and system logic design should stress an "open-endedness" about the possibilities of providing tutorial interaction between the machine system and the user. The near future will bring considerable progress in this area. Although there are no revolutionary ideas on the horizon, the increased labeling and cross referencing made possible by mechanical stores should allow for fruitful experimentation. In addition to labeling schemes, optimum search strategies must be sought. It is not enough that machines and men interact so as to converge upon a mutual understanding of what the system contains and what the searcher seeks. They must converge rapidly and economically, without tedious exploration. If these ends can be attained, increased indexing will be done by the user in the course of use, and not by increasing the indexing staff.

Further study on the possible dialogues between the user and the automated catalog is needed. These dialogues must be developed in considerable detail, flow charted, and then used as the basis for computer programming. Several novel techniques proposed within the last few years certainly hold sufficient promise to justify inclusion in some form in the proposed system. The use of citation indexing and the maintenance of user records represent particularly valuable sources for different approaches to information. A thorough study should be made of present descriptive cataloging rules to determine whether they should be modified in order to build maximum flexibility of use into the system and permit effective exploitation of the capability provided by automation and if so, how?

Communication Technology. It is unlikely that the needs of research libraries could appreciably affect developments in the area of communication technology. However, the reverse is not true. Centralized research libraries which provide service over telephone or telegraph lines are feasible now and are certain to become more attractive economically as communication technology advances. It is thus likely that, even at the outset, data communication facilities will be used, but quite sparingly, say, for priority information to the more important cooperating libraries. The research and development needs are minor, with the possible exception of terminal sets. The following questions need to be answered prior to large-scale use: To what extent is there a need for an extensive communication net? Will there be user acceptance? At what rate will the cost-performance ratio improve? Can orderly growth in facilities be achieved?

Textual Data Conversion Methods. Large-scale conversion of textual information into machine-readable form currently appears to cost on the order of one cent per word and proceeds at human typing rates. The development of flexible, multifont print readers, promising character-by-character conversion at several hundred characters per second, is being fostered by several Government agencies. However, the automatic conversion of files, such as the National Union Catalog, which contain considerable heterogeneous symbolism, is not likely in the near future. New methods for accomplishing catalog conversion should be given emphasis.

# *Consequences of Automation*

## *Impact on the National Library System*

Information Flow. It is evident that automation of the Library of Congress alone would not be a complete solution of all national research library problems. On the one hand, the Library of Congress does not hold everything published. Other libraries (the National Agricultural Library, the National Library of Medicine, etc.) have more complete collections in certain topical areas, and indexes to their holdings are better organized or are in more descriptive depth for the specialist's use. On the other hand, other libraries have come to depend upon the Library of Congress for assistance in carrying out their functions. The Library of Congress provides catalogs, lends books, distributes cards, and offers varied reference, bibliographic, and consultant services.

Thus it is desirable to conceive of a library network, a national research library system, incorporating the telecommunications necessary to accommodate the flow of information to all its branches. Technical developments in the last year or so have greatly improved this situation, and there is no doubt that intercommunication of the information considered here is technically feasible and economically reasonable.

At the present time, all information that is useful for the purposes intended to be served by the large research libraries of the United States is not stored in those libraries. For example, in the field of scientific research, technical reports and memoranda often are never brought into the library system or are brought in too late to be of maximum value. In this respect the importance of information centers (such as exist for handling technical literature dealing, for example, with nuclear science, aerospace, or military research) should be recognized and the relation of those centers to a national research library system examined. In many fields, especially those in which knowledge is accumulating most rapidy, the centers are serving a user need that libraries have neglected. It is not apparent, however, that information centers are able to do much more than provide an alerting function. If this is correct, then the research library system can assume the very important tasks of stimulating a fresh and useful flow from these newer information sources, and it can also provide orderly bibliographic control over their products so that they have long-term use.

Interlibrary Cooperation. Cooperation among libraries exists in acquisition, cataloging, particular bibliographic projects, library lending, and in many other areas. This cooperation is an attempt to make maximum use of limited resources. Libraries cooperate to improve service and in the interest of the common good, but the test to justify cooperation is usually an economic one. Automation will permit a change in outlook.

The centralized cataloging possibilities implicit in an automated library system could be realized only through a major reorganization of effort in individual libraries. Even acquisition attitudes will have to change as information on holdings is more and more rapidly exchanged and reproduction simplified. The benefits of integrating resources will, in general, be more difficult to ignore. The argument that libraries cannot effectively use the fruits of cataloging efforts performed elsewhere because the results do not meet their own local requirements is far less persuasive in an automatic system. This is so because of the possibility of the central cataloging function accommodating itself to a variety of user needs and perhaps performing special editing processes by machine in order to deliver a tailormade product to certain libraries having such a requirement.

Numerous advantages will accrue to individual libraries because of their having direct access to the National Union Catalog. If their own systems are automated, and they have provided for maintaining a use-frequency history within their collection, they can rely more heavily on the Library of Congress for the less frequently used books in order to keep valuable space in local libraries allocated to the active part of the collection.

## *Impact on the Library User*

The library-user relationship will be significantly affected, although all possible consequences cannot be foreseen. Clearly, services will become much more closely geared to the user's needs. Emphasis will be placed on saving the user time in finding and examining items of interest to him. The user will be assured a more comprehensive search and more selective service. As the users acquire experience with consoles, and as the system and its programs are designed to take into account the users' annotations, comments, and the use history of a variety of users and a variety of works, the system itself will become more responsive. In short, what will be accomplished is a much closer intellectual interaction between the user and the library. Only with this kind of improved interaction can the task of communicating recorded knowledge be performed with a high degree of effectiveness, so that major portions of it are not lost to posterity.

In a certain sense, the automatic library described here can be viewed as both a learning and a teaching machine, and in this way would have a great influence in enhancing the level of our culture.

## *Effect on Availability of Scientific and Technical Information*

The nation's problem of scientific and technical information has been of growing concern to the defense establishment and to the Congress.[1] Most recently the President's Science Advisory Committee published a report entitled *Science, Government, and Information,* in which the fragmentation of scientific literature by speciality is pointed out as demanding some counteracting consolidating force in communications:

"Though a scientist chooses to narrow his speciality, science itself creates an ever-increasing number of potential points of contact between the scientist's narrow specialty and the surrounding fields. As time goes on, successful pursuit of a narrow specialty requires effective contact with more and more diverse parts of the literature."

To see what the effect of automation might be, let us examine the three levels through which information passes in the publication process.

First, at the generation or first-publication level, individual workers or small groups record information that they have assembled by creative thought, experiment, or literature research in a publishable article, paper, or communication. Examples of this first-publication level are individual journal articles, doctoral dissertations, contract-reporting documents, industrial research reports, and so forth.

Without question, the input of knowledge to the generation level must exist as a diverse and spontaneous activity, and it is reasonable to expect that the publishing effort for this level will similarly remain diverse and essentially uncoordinated, even though the basic mechanisms of such publication may change appreciably over the years.

The second level is the book. Here, some major topic is comprehensively brought into a cohesively organized form summarizing and referring to prior published knowledge. At the book level, information, usually 2 to 10 years old, is assembled into a packaged form generally having some unifying theme. Periodicals are increasingly devoting entire issues to the treatment of all aspects of some topic. Such special issues, being equivalent to books but unlike them in their multiple authorship, suggest a decline of books as research sources. A comprehensive, mechanized reference and indexing system should allow such special-issue documents to be rapidly compiled and narrowly focused. One can also foresee the pruning of the now too prevalent weaker articles in such collections and, eventually, even the pruning of weaker portions of articles. The efforts of the Atomic Energy Commission, National Institutes of Health, Defense Documentation Center (formerly Armed Services Technical Information Agency), National Science Foundation, National Aeronautics and Space Administration, and others will assist and encourage this aspect of the organization and consolidation of knowledge.

The third level, the preparation of encyclopedias and treatises is declining, relative to growth of publication. Private enterprise has certainly provided an adequate encyclopedic effort when "encyclopedia" is considered in its popular sense. However, with a few

[1] See, for example, the following reports:

U.S. *Congress. Senate. Committee on Government Operation.* Documentation, indexing, and retrieval of scientific information; a report. Washington, U.S. Govt. Print. Off., 1960. 283 p. (86th Cong., 2d sess. Senate. Document no. 113) Addendum to this report issued as 87th Cong., 1st sess. Senate. Document no. 15.

U.S. *Congress. House. Committee on Science and Astronautics.* Dissemination of scientific information. Washington, U.S. Govt. Print. Off., 1959. 12 p. (86th Cong., 1st sess. House. Report no. 1179)

exceptions, the job of summarizing knowledge in particular subject areas in a systematic way has not been successfully attacked. It is this aspect of the organization of knowledge that could be undertaken under the aegis of an automated research library system. The consoles and machine-aided search strategies would greatly facilitate this work.

## *Organization of the Collections*

The Library of Congress and most research libraries shelve books in a classified array. The purpose is to place books on the same subjects together for stack control, delivery, special collections, browsing, etc. This mode of shelving has been questioned for a number of reasons: It is wasteful of space. Books are not usually written on a single subject, yet must be placed in only a single location. The number of books on a given subject in a large research library defeats browsing. It has even been said that browsing is an admission of defeat in indexing. Against this background of readiness for change, automation might be at least as welcome as it is disturbing.

Certainly, mechanization will impose a change, because a large research library, automatically controlled to insure the effective location and delivery of its holdings, cannot permit open-shelf operation. Browsing can now be freed of its dependence on classification, from which most of its ills stem, and given a functional orientation. Special browsing collections can be assembled of most frequently used materials or materials of established value for research in given subject fields. These collections can be seeded with analogous or related materials which the librarians believe to be of value and can be weeded, updated, and organized to make the browsing yield more effective and valuable.

A capability for browsing by use of the console should be of even greater significance. The opportunity to examine statistical data on the number of entries in a biliography, the bibliography itself, and then selected pages of particular items which may include title pages, tables of contents, and indexes, all on a successive rapid response basis and coupled with subject access to whatever depth economics permits, will provide a far more flexible intellectual interaction between the user and the collection than occurs in wandering through the stacks as though one were shopping in a supermarket.

## *Measurement and Feedback of Library Effectiveness*

It is a curious and unfortunate fact that libraries of today have virtually no way of knowing how well they are performing from the point of view of what the user ought to be getting from the system. In general, there is no way of knowing how much information responsive to a subject-oriented request is not found. Possibly one of the most important consequences of automation will be to provide a capability for maintaining use history and for implementing measurements on a sampling basis in order to install a good system of quality control in the library's operations. With such quality control measurements, the possibility for improvement-feedback then exists. It is unlikely that industry could operate today without an adequate system of quality control over its products. Why then should not libraries, with perhaps far more important products, stand to benefit by such a philosophy?

## *Concluding Comment*

Certain trends toward decentralization and specialization of collections in both the arts and sciences are partial evidence of the difficulties which research libraries are encountering in rendering adequate service. The most serious problem is that of characterizing documents to facilitate their retrieval and the retrieval of their contents within the bounds of a manual system. Many of the mechanical limitations could be overcome by the application of current or imminent technical advances. One can also be optimistic about increasing the sophistication of library use and, more importantly, the knowledge of how to use them by the advent of machines.

This survey has been very conservative in its appreciation of technology and science applicable to document and information retrieval, but some of the team have more optimism about future developments than can be documented by today's equipment and methodology. The need for automation in libraries will create advances in these areas.

It is a general conclusion of this survey that research libraries are lagging behind in the application of existing technology to their information handling problems, principally because of insufficient funds. Based on its investigation, the survey team asserts that existing

and imminent technology can bring about major improvements in the service of the Library of Congress, and that an automated system accomplishing this can be made operable within the next few years. The survey team has therefore recommended that immediate efforts be made to automate research library functions. It is also strongly recommended that the Library of Congress, because of its central role in the Nation's library system, take the lead in the automation venture.

# APPENDIX

715-690—63——3

# A COST ANALYSIS OF AN AUTOMATED SYSTEM FOR THE LIBRARY OF CONGRESS

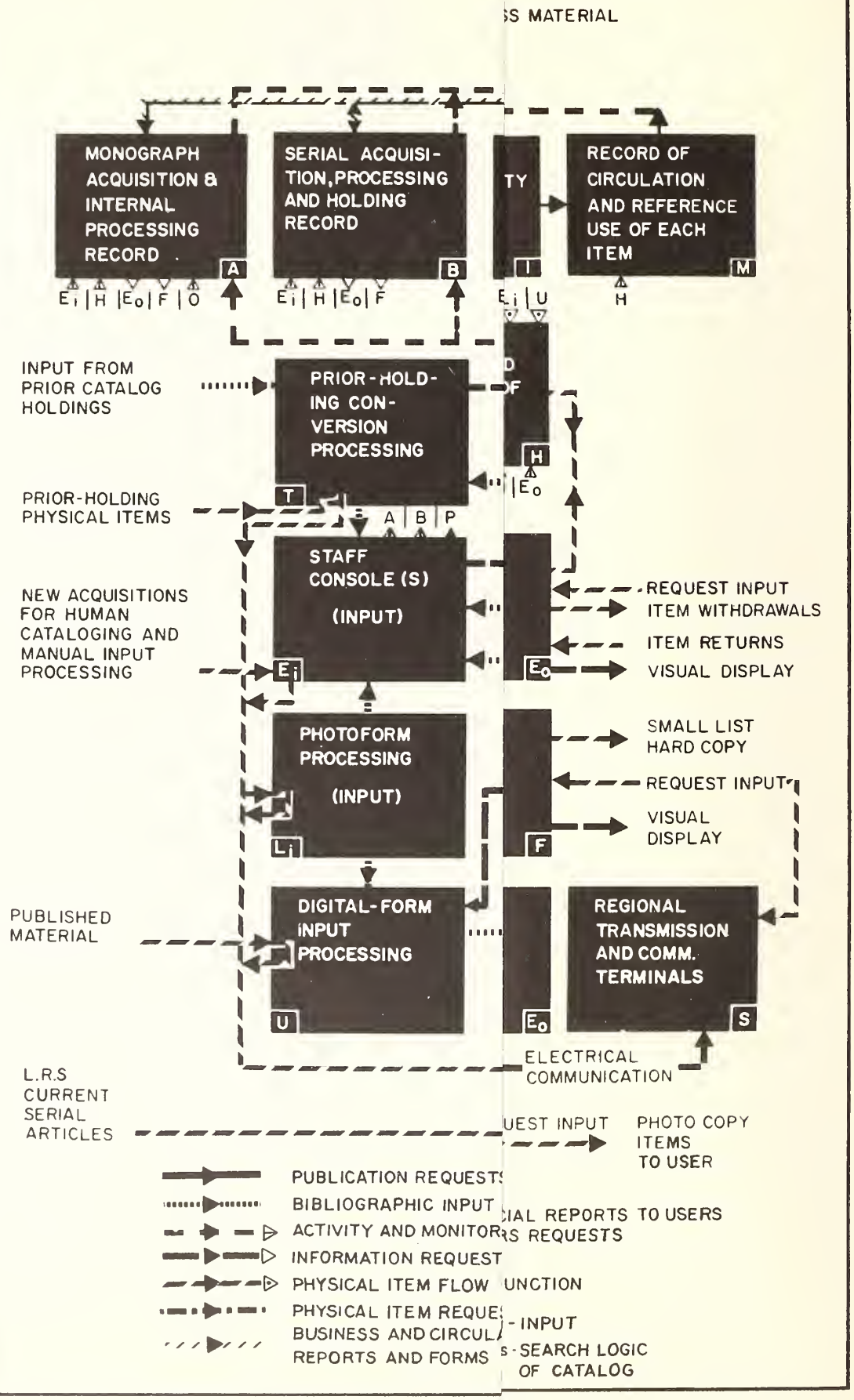
Prepared under the direction of the survey team
by Herbert T. Spiro and Allan D. Kotin
of the
Planning Research Corporation, Los Angeles, California,
August 1962

# *Foreword*

The following study was made under the direction of the survey team to determine the applicability of automation and its costs. It was recognized that even a preliminary assessment of the money and effort involved in automating portions of the Library of Congress would require that the functions of the Library be analyzed and a system proposed for handling those functions found amenable to automatic execution.

Accordingly, an analysis and system design were undertaken, but with the restricted objective of developing some preliminary tools to aid in constructing the basis for a research and development program. The survey team wishes to point out that the results, as reported here, should not be interpreted as the outcome of such a research and development program, nor should they be interpreted as an endorsement of any particular technique or device or as a set of inflexible recommendations. Based on the experience of librarians and computer experts, goals were chosen and equipment for realizing them suggested as technically and operationally feasible by 1972.

It is the opinion of the survey team that trends in cost and performance in the computer field are now sufficiently well established so that the estimates arrived at will retain validity, even though specific details of the system may change extensively.


SS MATERIAL
MONOGRAPH ACQUISITION & INTERNAL PROCESSING RECORD
A
Ei | H | Eo | F | O
SERIAL ACQUISITION, PROCESSING AND HOLDING RECORD
B
Ei | H | Eo | F
TY
I
Ei | U
RECORD OF CIRCULATION AND REFERENCE USE OF EACH ITEM
M
H
INPUT FROM PRIOR CATALOG HOLDINGS
PRIOR-HOLDING CONVERSION PROCESSING
T
A | B | P
H
Eo
PRIOR-HOLDING PHYSICAL ITEMS
STAFF CONSOLE (S) (INPUT)
Ei
NEW ACQUISITIONS FOR HUMAN CATALOGING AND MANUAL INPUT PROCESSING
REQUEST INPUT
ITEM WITHDRAWALS
ITEM RETURNS
Eo
VISUAL DISPLAY
PHOTOFORM PROCESSING (INPUT)
Li
SMALL LIST HARD COPY
REQUEST INPUT
VISUAL DISPLAY
F
PUBLISHED MATERIAL
DIGITAL-FORM INPUT PROCESSING
U
Eo
REGIONAL TRANSMISSION AND COMM. TERMINALS
S
ELECTRICAL COMMUNICATION
L.R.S CURRENT SERIAL ARTICLES
UEST INPUT
PHOTO COPY ITEMS TO USER
PUBLICATION REQUESTS
BIBLIOGRAPHIC INPUT
ACTIVITY AND MONITOR
INFORMATION REQUEST
PHYSICAL ITEM FLOW
PHYSICAL ITEM REQUE
BUSINESS AND CIRCULA
REPORTS AND FORMS
IAL REPORTS TO USERS
S REQUESTS
UNCTION
- INPUT
S - SEARCH LOGIC OF CATALOG

SECTION I

# *Background and Summary*

## *Background*

In order to assess the economic feasibility of library automation, a study of the costs of hardware, software, salaries, and maintenance of a hypothetical automated Library of Congress was undertaken. The year 1972 was selected as the date the system is to be fully operational, and projections of Library operations expected to be directly affected by automation were made. The automated system used as the basis of this study is one of several possible configurations (see Exhibit 1); it should be interpreted as a model and not as the recommended system. The system does, however, account for complete or partial automation of all the functions which the survey team considered necessary for the basic operation of an automated research library. (For a list of these functions, see Exhibit 2.)

The report is thus function-oriented. Although it would have been possible to describe existing or proposed hardware and then specify in what manner this equipment could aid in Library operations, this approach was considered unrealistic. The organization of this report essentially follows the methodology of the cost study: present Library activities are analyzed; interrelationships and duplications due to the constraints of manual operation are noted, in addition to desirable expansion of services; a list of Library functions is presented; these functions are then described operationally; workload statistics are developed for the various operations comprising a function; data are projected for 1972; specifications for hardware to perform the various functions are described; and, finally, costs are developed. In general, the analysis has been kept simple so that this functional orientation is always apparent. In the various possible systems which further studies might propose, the hardware complexes might differ, but these functions would remain essentially unaltered.

EXHIBIT 2.—FUNCTIONS OF AN AUTOMATED LIBRARY SYSTEM

| *Function Code* | *Function* [1] |
|---|---|
| A..... | Recording monograph acquisition and processing. |
| B..... | Recording serial acquisition, processing, and holdings. |
| C..... | Maintaining cataloging authority files. |
| D..... | Catalog use. |
| E..... | File access for staff. |
| F..... | File access for readers. |
| G..... | Control of Library holdings. |
| H..... | Physical storage of items. |
| I..... | Circulation control. |
| J..... | Editing for input and output. |
| K..... | Publication of bibliographic materials. |
| L..... | Production of photocopy. |
| M..... | Monitoring use of the collection. |
| N..... | Monitoring catalog use and response. |
| O..... | Business data processing. |
| P..... | Maintaining exchange and gift records. |
| Q..... | Scheduling Library operations. |
| R..... | Providing special graphic retrieval. |
| S..... | Communication with other institutions. |
| T..... | Converting existing files to machine-readable form. |
| U..... | Processing materials received in digital form. |
| V..... | Controlling intrasystem communication. |

[1] Terminology for the various functions may differ slightly on the Functional Block Diagram (Exhibit 1); the function codes are identical.

Functional analysis necessarily cuts across administrative lines. Since automation is not necessarily suitable for all present operations of the Library, each division of the Library was studied to determine which of its operations could be assigned to the functions developed. As a result, certain operations performed by the following divisions were identified as the primary candidates for extensive automation:

*Processing Department*

Catalog Maintenance Division
Decimal Classification Office
Descriptive Cataloging Division
Exchange and Gift Division
Order Division

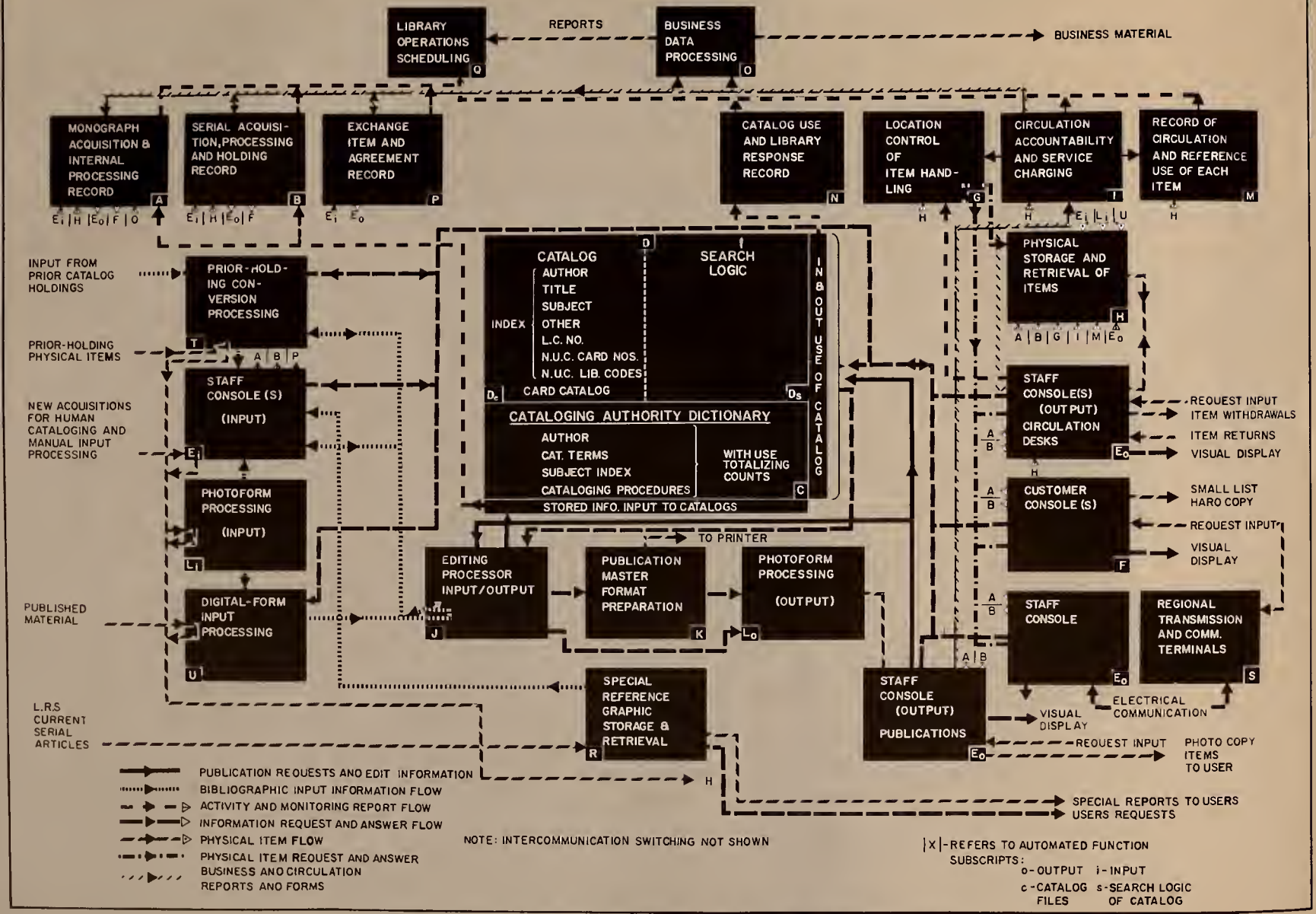



EXHIBIT 1—FUNCTIONAL BLOCK DIAGRAM OF AN AUTOMATED LIBRARY SYSTEM

Serial Record Division
Subject Cataloging Division
Union Cataloging Division
*Reference Department*
Loan Division
Serial Division
Stack and Reader Division

In addition, partial automation could be anticipated for seven other divisions of the Reference Department, as well as certain activities of the Law Library and the Legislative Reference Service. In Section II, the activities of the Library are reviewed, and reasons for including or excluding them in the automated Library are given.

## *Summary of Findings*

The costs of the automated library system include initial expenditures for the acquisition of the system and recurring expenditures for its operation and maintenance. Exhibit 3 presents a summary, in constant dollars, showing a total cost of $28 million for procuring the system. Of this amount, $16.7 million is for the acquisition of hardware and $11.6 million for software. Adding the cost of full conversion of files to machine-readable form ($3.3 million) brings the total initial cost to approximately $31.6 million. Annual operating costs amount to about $770,000, of which $452,000 is for salaries of programmers and equipment operators, and $318,000 for maintenance and use of consumable materials. (The derivation of the above figures may be found in Sections II to V and Section VII.)

The cost of converting existing catalogs and files to machine-readable form is of special interest. For the system described, some records, e.g., acquisition and circulation records, would have to be completely converted before the system could operate, while other files, e.g., the card catalog, could be partially converted. The results of the cost analysis show that the cost of converting all relevant Library files (about $3.3 million) is only about three times as much as the cost of converting the minimum required files ($1.2 million) although the number of words converted is estimated at 700 million in the first case and only 53 million in the latter, a volume $\frac{1}{13}$ as large. Full conversion, therefore, appears advisable.

The projection of the costs of the present manual system provides a yardstock for comparing the costs of the automated system. The initial costs of the automated system are almost $30 million; the annual operating cost of $4.5 million is about $0.5 million under expenditures projected for the manual system in 1972. It is of interest to contrast these estimates with total obligations incurred by the Library of Congress in recent years. These amount to $19.1 million in fiscal year 1959, $20.3 million in 1960, and $22.3 million in 1961. Assuming an average increase in obligations of $1.5 million per year, the budget for 1972 would be $38.8 million. Total obligations incurred between 1962 and 1972 without an automated system are estimated at $320.5 million; with an automated system, $352.2 million.

EXHIBIT 3.—PROCUREMENT COSTS OF AN AUTOMATED SYSTEM

| Hardware description | Estimated cost | | |
|---|---|---|---|
| | Hardware | Software | Total |
| Internal processing group | $744,000 | $671,000 | $1,415,000 |
| Circulation and retrieval group | 969,000 | 786,000 | 1,755,000 |
| Central catalog group | 5,000,000 | 5,000,000 | 10,000,000 |
| Consoles | 5,440,000 | (1) | 5,440,000 |
| Editing processor | 3,363,000 | 3,000,000 | 6,363,000 |
| Photocomposer | 500,000 | (2) | 500,000 |
| Business data and scheduling | 120,000 | 120,000 | 240,000 |
| Special reference graphic storage | 339,000 | 25,000 | 364,000 |
| Miscellaneous general hardware | 208,000 | (3) | 208,000 |
| Detailed system engineering design | (3) | 2,000,000 | 2,000,000 |
| Total | 16,683,000 | 11,602,000 | 28,285,000 |

[1] The cost of programming the set subroutines associated with console operation is included in the purchase cost of console hardware.

[2] This function will be controlled by the editing processor and all relevant software costs are attributed to the editing processor.

[3] Not applicable.

In order to view the costs of an automated Library of Congress in proper perspective, projections to 1972 of certain costs of the manual system were undertaken. This analysis indicated that, after the initial cost of procuring the system, the following comparisons of operations show that the salary costs are actually slightly lower in the automated system.

| Service group | Projected salary costs for 1972 | |
|---|---|---|
| | Manual system | Automated system |
| Cataloging and catalog maintenance | $2, 375, 000 | $2, 163, 400 |
| Acquisition and internal processing | 1, 297, 400 | 1, 013, 100 |
| Circulation, retrieval, and storage | 1, 313, 310 | 992, 400 |
| Total | 4, 985, 710 | 4, 168, 900 |

In the long run the costs of Library operations are dependent on the volume of activities. For this reason, particular emphasis was placed throughout this study on the selection, analysis, and projection of key statistics. The above costs are thus derived from an analysis of the Library workload for 1972. Three of the key estimates for developing hardware specifications are the following:

| | *Workload for 1972* |
|---|---|
| New titles cataloged | 120, 000 |
| Volumes retrieved from the stacks | 1, 100, 000 |
| Items recommended for addition to the collection | 210, 000 |

The data on which these and similar projections were based are presented in Section IV of this report.

The analysis of functions to be performed by the system and a projection of the volume of operations provided a basis for developing requirements for nine basic hardware groups. At the core of the system are three large digital storage files: the automated catalog and two smaller files for recording all processing and circulation operations. These files have the following characteristics:

| | Capacity (bits) | Expected frequency of access (use per minute) |
|---|---|---|
| Processing group | $9 \times 10^8$ | 22 |
| Central catalog group | $5 \times 10^{10}$ | 170 |
| Circulation and retrieval group | $1 \times 10^9$ | 39 |

The size of the central catalog in the above table is its initial size (assuming full conversion) in 1972; no provision is made for additional indexes or for future growth. Obviously file capacity on the order of $10^{11}$ to $10^{12}$ bits is required and, as shown in Exhibit 3, $10 million has been allocated for the hardware and software for this catalog.

In addition to these three files, six additional hardware complexes are defined and costed in this survey. Closely related to the automated catalog is a large computer system, called the editing processor, whose function is to edit, arrange, and transfer all inputs to the automated catalog. In addition, it will be designed to generate automatically demand bibliographies, special lists, and even complete catalogs in digital form. Another hardware unit, the photocomposer, converts this digital output to high quality masters which can then be used for the actual publication of bibliographies, lists, and catalogs for human consumption.

The essential link between readers, staff members, and these automated storage files would be provided by means of a console. More than 200 consoles, located throughout the Library, provide instant access to any storage file through the console keyboard and a display print facility. Other consoles may be distributed throughout the country and linked ultimately to the automated catalog by means of a regional communications network.

Further details on the above hardware groups and the groups not discussed here may be found in Section V. The hardware discussed in this section is either "on the shelf" or within the state-of-the-art in advanced developmental phases.

## *Conclusion*

This study has aimed at developing the best available data to assess the economic feasibility of automating certain Library processes. Most of the justification has been based on a comparison of costs between the manual system and the automated system; no attempt has been made, however, to determine what it would cost to provide by manual methods all the services offered by the automated system. In other words, even if the costs were the same, the benefits are demonstrably greater in the automated system, since it would provide a wide variety of new services and a greater refinement of existing services. In most cases these

715-690—64——4

services can be considered as cost-free byproducts of automation. Even those services which require some incremental salary expenditures, however, are practical only by virtue of the economies of automation.

The nature of the automated catalog makes possible a number of benefits not presently available. The provisions for searching the catalog by date, language, and other criteria would greatly enhance most types of reference use as well as the production of bibliographies. The rapid and inexpensive production of bibliographies and lists for special subjects, authors, publishers, etc. would greatly enrich the service provided to research workers. The production of such bibliographies at the present time requires a lengthy and costly effort, and the Library is unable to undertake this work in response to the needs of individual scholars.

One of the greatest services of the automated system is the aid which could be given to libraries throughout the country in organizing and cataloging their collections. The automated catalog could incorporate the National Union Catalog and could be indexed by the location of individual titles in cooperating libraries. The full impact of the significance of this service cannot readily be measured at this point, but, as the pressures on educational institutions increase, cooperation between research libraries is increasingly important. As the library community continues to standardize cataloging rules, the indexing provisions of the automated catalog would effectively make it a truly national catalog providing integrated files of virtually all published materials held by research libraries. Of perhaps greater significance is the possible use of this catalog for automatically generating catalogs and accessions lists in either printed or digital form for cooperating libraries. The production of these publications would require relatively simple programming of the editing processor and would be well within the capabilities of the equipment required simply to meet the internal needs of the Library of Congress.

The digital storage and automated control of all processing records would undoubtedly produce substantial savings in time and effort by virtue of the ease with which claims for missing serial issues or other items could be generated. Full integration of exchange and gift activities with other acquisition and processing records would permit more efficient use of present exchange sources for the benefit of the Library of Congress and for the enrichment of the numerous libraries throughout the world with which the Library maintains exchange programs.

An additional benefit, which would be difficult to measure in terms of cost, is the virtual guarantee of file integrity that digital storage permits. The unit record would no longer be a catalog card, which can be removed, misfiled, or tampered with; instead, it would be a digital record filed automatically and therefore inviolate. With such a file, much of the double checking and special searching required in both reference and cataloging operations would be eliminated.

These benefits are of little use, however, unless one can readily obtain the materials selected after consultation of these automated files. The present system maintains records for materials circulated outside the Library, but, for the most part, no records are maintained of items used by readers within the Library. The replacement of the present call slip with automatically typed and magnetically readable labels for items requested from the stacks would permit greater control over the collection and the precise determination of items not in their regular shelf locations. The use records thus generated would also permit relocation of materials to maximize retrieval time and minimize storage space.

The automated system discussed in this report is described in a rough and preliminary manner. Development of more detailed specifications for the system would inevitably bring to light other major benefits to the Library, the Library user, and to other libraries, the marginal costs of which would be negligible. Only a complete and detailed system design can be expected to describe all, or even most, of the possible combinations and operations of the proposed system. For this reason this study has largely described a system which can perform the operations now performed manually and at an economic advantage; it must be kept in mind, however, that in addition to matching the manual system the automated system would provide additional cost-free benefits, of which the above are only the most obvious and readily identifiable.

Each of the sections which follow deals with a particular subject and is organized as an independent paper. Readers interested in particular areas can concentrate on the relevant sections without exploring peripheral issues.

# *Identification of Major Library Activities*

## *Summary*

This section concentrates on identifying those activities which are directly related to the central bibliographic record and which therefore would be altered by automation. The following divisions of the Processing Department will be directly affected by automation: the Decimal Classification Office and the Catalog Maintenance, Descriptive Cataloging, Exchange and Gift, Order, Serial Record, Subject Cataloging, and Union Cataloging Divisions. Within the Reference Department operations of the Loan, Serial, and Stack and Reader Divisions will be altered by the installation of an automated system. Other divisions in the Library will be affected to the extent that their operations are changed by the overall improvement of library services, such as increased speed of access to automated bibliographic records and catalogs, increased speed of processing incoming materials, and the production of bibliographies and lists through automated equipment.

## *Activities Considered for Cost Analysis*

The Library of Congress performs many functions which make it in many respects a unique institution. Among these functions are the regulatory duties of the Copyright Office and cultural activities—poetry readings, lectures, concerts, and exhibits—which, by their very nature, are not within the scope of this study. A general rule used to include or exclude a library activity was the extent to which it was bibliographic or nonbibliographic. In the following pages a more precise enumeration of criteria for inclusion will be given, along with detailed analyses of operations amendable to automation.

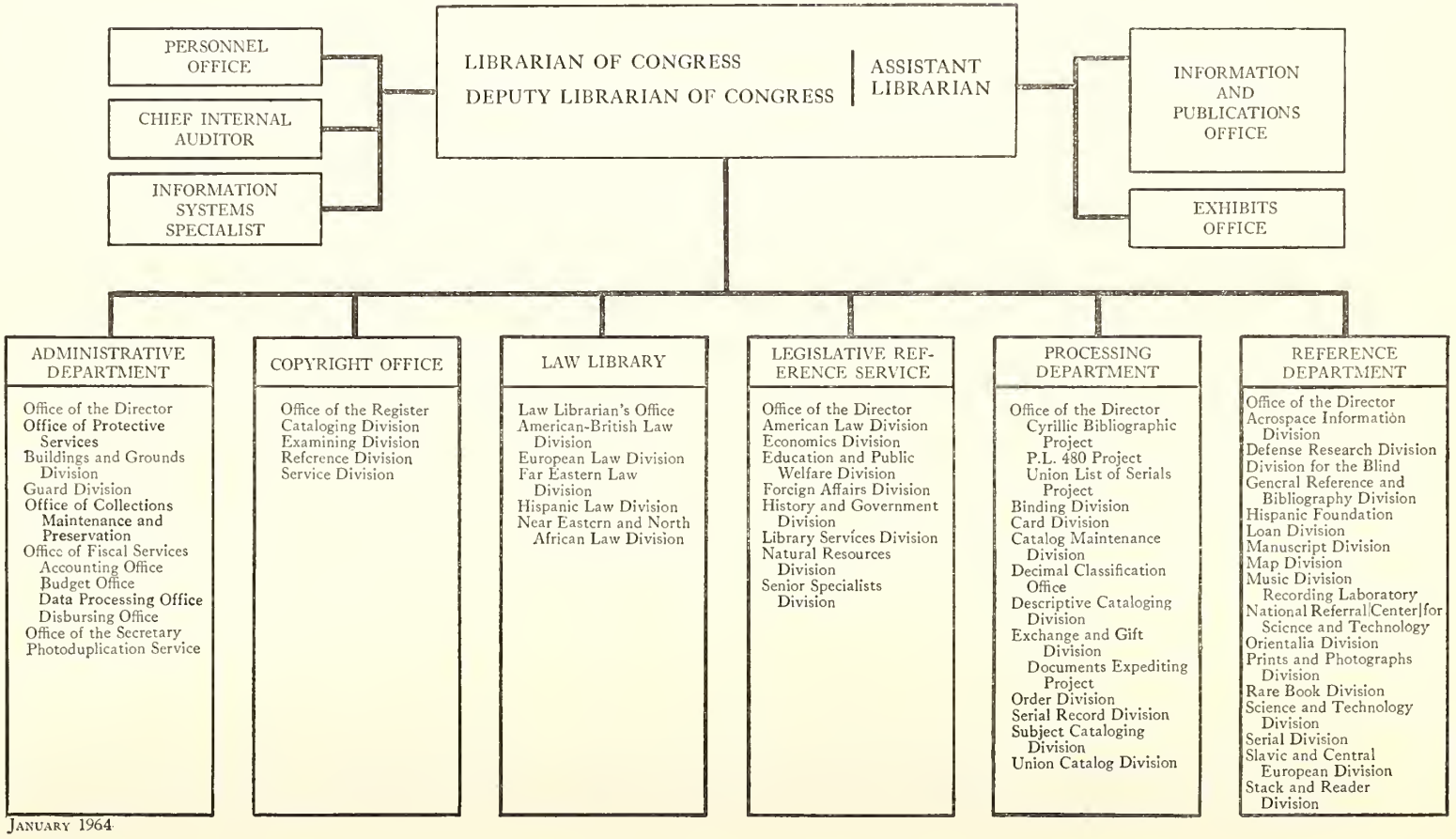
A facsimile of the most recent Library organization chart is presented in Exhibit 4. The Office of the Librarian and the five related offices at the top of the chart, as well as the Administrative Department, and the Copyright Office, were eliminated from consideration. This does not imply that some operations within these units could not benefit from automation but rather that their underlying function is administrative or regulatory rather than bibliographical.

The Law Library and the Legislative Reference Service were largely excluded, except for the provision of equipment to permit their staff members to communicate with and interrogate all mechanized Library files. The former was excluded because its collection, kept physically separate from the general Library, is focused largely on units of information peculiar to legal source material, such as legal indexes, looseleaf services, and records of legislative proceedings, and hence not too easily incorporated with the bibliographic units contemplated for the general automated Library. The Legislative Reference Service, which serves primarily as a research staff for the benefit of Members of Congress, provides bibliographic and research service of a highly specialized, individualistic, and somewhat unpredictable nature. Both Law Library and LRS staff members will benefit from automation when their research activities require use of the central catalog and associated files and from the expected increase in speed of processing incoming materials. In addition, the special reference files developed and maintained by the Library Services Division of LRS are considered to be suitable for automation by special fast-access graphic storage, and this aspect of LRS operations will be given further attention.

The Processing Department and the Reference Department, each of which includes several divisions, are most directly affected by automation although, again, certain divisions and activities are either excluded entirely or treated briefly. The remainder of this section is largely concerned with a discussion of operations in these departments which are suitable for automation. (See Exhibit 5.)

## EXHIBIT 4

# LIBRARY OF CONGRESS
## Organization Chart




JANUARY 1964

EXHIBIT 5.—SUMMARY OF DEPARTMENTS WITH RESPECT TO COST ANALYSIS

| *Department* | *Treatment in this report* |
|---|---|
| Office of the Librarian of Congress: | Excluded from analysis except for automation of Library operational monitoring functions (indirect). |
| Office of the Deputy Librarian. | |
| Office of the Assistant Librarian. | |
| Personnel Office. | |
| Chief Internal Auditor. | |
| Information Systems Specialist. | |
| Administrative Department | Excluded from analysis. |
| Copyright Office | Excluded from analysis. |
| Law Library | Included in analysis of procurement costs of catalog access equipment. |
| Legislative Reference Service | Included in analysis of procurement costs of catalog access equipment and special reference graphic storage equipment. |
| Processing Department | Included in the analyses of procurement costs and salary costs.[1] |
| Reference Department | Included in the analyses of procurement costs and salary costs.[2] |

[1] Certain divisions in this department are excluded (see Exhibit 6).

[2] Certain divisions in this department are excluded (see Exhibit 7).

## *The Processing Department*

Exhibit 6 indicates the extent to which this report treats the Processing Department. Excluded activities include administrative functions; special projects listed under the Office of the Director (these would be affected by automation, but are special, noncontinuing endeavors); the Binding Division; and the Card Division.

The primary activity of the Binding Division is the monitoring and preparation of materials forwarded for binding. Much of this operation is manual but not subject to automation; for example, issue-by-issue collation of serials, checking volumes for correct placement of title pages and indexes, etc. The recordkeeping functions associated with binding are partially included in the automation of serial and monograph processing.

The exclusion of the Card Division from the cost analysis of a proposed automated Library requires explanation, particularly since many division operations appear to be well suited to mechanization. As described in a "Selective Outline of Functions in the Library of Congress," prepared by the Office of the Information Systems Specialist, division responsibilities include the following:

1. Estimates number of cards to be printed and assigns card numbers.
2. Maintains physical stock of printed cards.
3. Maintains catalog of printed cards filed by main entry.
4. Receives orders for printed cards by card number, author, subject, etc., and conducts required searching operation to fill orders.
5. Fills standing orders for cards by subject, for proofsheets (in complete or partial sets), and for series or parts of series.
6. Distributes certain bibliographies and catalogs prepared in the Library.
7. Provides proper business operation for billing and accounting.

As may be seen from this list, many of the operations carried on by the Card Division are specifically oriented to provide catalog cards and proofsheets to other libraries. In this sense, the Card Division may be considered as a business operating within the Li-

EXHIBIT 6.—DIVISIONS OF THE PROCESSING DEPARTMENT

| Division | Costs of automation estimated | |
|---|---|---|
| | Yes | No |
| Office of the Director | ..... | X |
| Cyrillic Bibliographic Project | ..... | X |
| Union List of Serials Project | ..... | X |
| Binding Division | ..... | X |
| Card Division | ..... | X |
| Catalog Maintenance Division | X | ..... |
| Decimal Classification Office | X | ..... |
| Descriptive Cataloging Division | X | ..... |
| Exchange and Gift Division | X | ..... |
| Order Division | X | ..... |
| Serial Record Division | X | ..... |
| Subject Cataloging Division | X | ..... |
| Union Cataloging Division | X | ..... |

brary. The importance of these business operations are illustrated by the fact that in fiscal 1961 about 35.7 million catalog cards were sold for a dollar volume of $2,040,000.[1]

There are two major reasons for the exclusion of the Card Division from this analysis. It is anticipated that one of the consequences of automation will be the eventual elimination of the necessity for card catalogs in the Library of Congress. Since card catalogs would be replaced by automated electronic catalogs, the production of printed catalog cards would either be automated as a part of the automated publication function or would possibly become a purely internal activity of the Card Division. In either case, there would cease to be any integral connection between the Library's internal operations and the business operations of the Card Division, except that the Library catalogers would continue to prepare data for card copy.

The second reason for exclusion of the Card Division relates to the type of hardware needed for automating its operations. Although the automation of the general Library catalogs would permit Card Division searchers to identify more rapidly the cards ordered, there would still be the problem of reproducing these cards in large numbers. The extent to which the automated publication function of the Library could handle the constant reprinting of cards, in addition to the other publication loads, cannot be determined without a more precise systems analysis of Card Division operations and further specification of the automated publication equipment. Even if the automated searching and printing of cards ordered were possible with equipment provided for other functions of the Library, there would still remain the major problem of matching individual printed cards with specific orders. The huge task of assembling orders for shipment to their buyers probably could be solved at least in part by automation. The specific nature of the system required for the Card Division would require specialized study of its operations and would almost inevitably involve electronic equipment for use only within this division and not of particular benefit to other operations of the Library.

[1] Data available since the preparation of this report indicate that, in fiscal 1962, 42.4 million cards were sold for $2,150,400. (The dollar volume of sales is not directly proportional to the number of cards sold, since the price of cards range from 1 to 13 cents; total sales are thus dependent on two variables, the number of cards sold and the type of cards sold.)

Unlike most divisions in the Library, the Card Division is self-supporting. The revenues from the sales of cards are returned to the United States Treasury and usually balance or surpass the expenses of the division. It is therefore reasonable to expect that automation of Card Division operations will be undertaken only if the costs incurred can be met by sales revenue. The primary purpose of this report is to estimate the procurement and operating costs of an automated system for the bibliographic operations and not for those business operations which are self-sustaining. In view of these considerations, further analysis of the Card Division was not deemed pertinent.

## *The Reference Department*

The extent to which this report considers the various divisions of the Reference Department is illustrated by Exhibit 7. The Manuscript Division, Map Division, Music Division, and Prints and Photographs Division were considered outside the scope of this report because their collections involve materials which differ markedly from the central library collection. The problems involved in maintaining collections of manuscripts, maps, photograph records, music scores, prints, and photographs are somewhat different from those characterizing the organization of monographic and serial literature. Many of the automated functions do not apply to these types of collections and could not readily be adapted without extensive modification.

The Orientalia Division presents problems because of the nature and diversity of the alphabets with which it deals. This, in conjunction with the relatively small collections involved, does not make it a candidate for immediate automation. There is, of course, the possibility that rapid technological progress will lower the cost of processing non-Roman alphabets to the point where such automation is practical even for limited collections, but the necessary technology is not sufficiently defined at this time to permit definitive assessment of its application.

The Division for the Blind was excluded because of the special nature of the bibliographic materials with which it deals, but it does deserve further careful consideration in the systems design studies. The Rare Book Division, General Reference and Bibliography Division, and Science and Technology Division are considered only to the extent that they require electronic equipment for access to the catalogs, since many of their operations will continue essentially unchanged

by automation. Although these divisions will be significantly affected by the greater refinement and ease of access resulting from automated cataloging, catalog access comprises only a small part of their total effort, and whatever savings are effected would probably be reflected more in increased and better service rather than in replacement of personnel.

Similar considerations apply to the Defense Research Division and the Aerospace Information Division. These divisions, although administered by the Library, are operations funded by the Department of Defense for the purpose of organizing and disseminating information. Certain procurement costs are included; salaries, however, are excluded. The Slavic and Central European Division and the Hispanic Foundation, again, are considered only with respect to the cost of supplying consoles for access to the automated catalogs and files.

The Loan Division, Stack and Reader Division, and Serial Division are all dealt with comprehensively in this report. These divisions are responsible for the circulation of material within or outside the confines of the Library. All will be directly affected by automated functions concerned with controlling item location and circulation.

Despite the elimination of a number of divisions from detailed consideration in this report, the survey team is confident that an aggressively pursued automation program at the Library of Congress, within the scope considered and proposed, will ultimately lead to the subsuming of many of these "excluded" functions within the automated structure. Many of these functions should be given careful consideration during the recommended systems design study.

EXHIBIT 7.—DIVISIONS OF THE REFERENCE DEPARTMENT

| Division | Costs of automation estimated | |
|---|---|---|
| | Yes | No |
| Office of the Director | ..... | X |
| Aerospace Information Division | X | ..... |
| Defense Research Division | X | ..... |
| Division for the Blind | ..... | X |
| General Reference and Bibliography Division | X | ..... |
| Hispanic Foundation | X | ..... |
| Loan Division | X | ..... |
| Manuscript Division | ..... | X |
| Map Divison | ..... | X |
| Music Division | ..... | X |
| Orientalia Division | ..... | X |
| Prints and Photographs Division | ..... | X |
| Rare Book Division | X | ..... |
| Science and Technology Division | X | ..... |
| Serial Division | X | ..... |
| Slavic and Central European Division | X | ..... |
| Stack and Reader Division | X | ..... |

# *Sketch of Functions To Be Performed by an Automated System*

## *Summary*

In this section the functions which the automated system should perform as a minimum are defined operationally without respect to the present administrative structure of the Library. In describing functions, every effort was made to make them consistent with each other and with the general system objectives. Exhibit 1, a functional block diagram of the proposed automated system, illustrates the interrelationships between these various functions.

Functions are grouped, for discussion, under the following general headings: acquisition and recording, cataloging, staff and reader file access, circulation and control, editing, administrative, and miscellaneous. In succeeding sections each function is described in greater detail with respect to magnitude of operations involved, hardware required, and probable cost in the automated system. Value judgments with respect to time phasing of these functions in the automated system have been indicated only in a very general way; essentially all functions are treated as deserving equal attention.

## *Acquisition and Recording Functions*

*Function A.* Recording Monograph Acquisition and Processing.

*Service provided:* Allows rapid identification of monographs that are either desired for the collections, on order, or likely to be acquired, and their status. Records status of all monographs throughout processing sequence, until final recording in the automated catalog.

The bibliographic description of monographic items would enter the acquisition and processing file at an earlier stage than under the manual system, since they would be listed when requested by a recommending officer in any of the various divisions. When actually placed on order or requested on exchange, monographs would receive an order number which would be posted in the file against the author and title entries. This function would embrace both the present order (purchase) and exchange records, but not the related financial operations which would be included in Function O. This file would also provide an index to all standing orders or exchanges.

As a monographic item proceeds through the stages of being recommended, ordered, received, and cataloged, each change in its status would be recorded and dated in this file. Thus the file also subsumes the present operations of the Process Information File. When the entire processing cycle is completed for a particular item, the entry is purged from the processing file and entered in the automated catalog.

As a part of this function, records would be kept for each item of the time spent in each stage of processing. There would be an automatic or semiautomatic program for flagging items overdue in either acquisition or internal processing; thus, a greatly improved control over this phase of Library operations would be possible.

*Function B.* Recording Serial Acquisition, Processing, and Holdings.

*Service provided:* Records serials on order or recommended for order and their status, as well as new serial titles being processed. Records Library holdings of all serial issues, binding records, and maintains a file of missing issues.

This function is somewhat parallel to Function A, the major difference being that the monograph file is constantly purged and therefore relatively constant in size, while the serial file will be continually growing.

Once a serial is ordered, the acquisition process is essentially continuous, and for that reason the file would not be purged.

One of the major uses of this file would be for posting receipt of each serial issue. An important requirement would be an automatic or semiautomatic program to note serials that have not been received, are due for binding, or are overdue from the bindery. Comprehensive lists of claims or overdue notices could be prepared automatically at regular intervals, again providing a control over the receipt and subsequent control of materials not possible with the manual system.

As a part of the posting operation, a routing slip would automatically be generated for each serial issue, indicating where it is to be sent. This is of particular importance only insofar as the present practice of routing current serial issues to various divisions and sections in the Library is maintained in the future.

## *Cataloging Functions*

*Function C.* Maintaining Catalog Authority Files.

*Service provided:* This file provides data on author entries, subject headings, and cataloging rules as established for use in the Library's files and catalogs. For each entry or heading there is an indication of the size of the file in the automated catalog. This file is basic to cataloging operations but should also be available to reference librarians, other staff members, and users who need assistance in determining proper entries for searching.

There is in the present system an authority file which is incorporated in the Official Catalog. In this file the following kinds of information may be found:

1. Authority cards showing the established forms of entries, the manner in which the authority was established (sources used for spelling and biographical data), and all necessary cross references.
2. Subject heading cards showing proper Library of Congress format and usage, cross references to preferred headings and references to related subject headings.

This automated catalog would incorporate the entire authority file for automated lookup to assist both catalogers and library users.[1] It should be noted that the authority file must be complete from the very beginning of the system, no matter what the size of the collection in the automated catalog may be. A partially complete authority file is of virtually no value in cataloging and of limited value in assisting library users.

As mentioned above, the file would automatically record each new assignment of a subject heading or author entry and would provide a current total of each file in the catalog. The usefulness of these counts applies in the cataloging and retrieval process and in other library operations as well.

The basic central file would contain, in machine-readable form, data equivalent to that on present LC cards. The file would be ordered and addressed by an item number, i.e. accession number or sequential conversion number. The item number might contain the designation of the year of publication or some other combination determined to be useful in searching and sorting. Contents of this portion of the automated system are essentially equivalent to the merged Official and Main Catalog of the Library (minus authority cards subsumed in the files related to Function C). In addition, it would have provision for incorporation of titles in the National Union Catalog not held by the Library, although these titles might possibly be separated in the files. This basic catalog would not require periodical merging if it were sequentially ordered.

The search logic portion of the automated catalog would govern search mode sequences and would provide for selection of proper sections of the files for searching. It is anticipated that use of the automated search logic would permit searches of a complexity not possible in the manual system. For example, the basic entry would contain digital field symbol codes which would identify all the various kinds of data contained on an LC card. This should permit limited searching by modes such as date and place of publication, language, number of pages, etc. Should any of these "new" search modes be widely used, the catalog could be programmed through Function J to generate additional indexes.

The inclusion of the NUC symbols for cooperating libraries allows the possibility of retrieving titles by location or indicating locations of a particular title. This means that it would be possible to generate a catalog of all the titles listed for a particular location, i.e. library, and to produce up-to-date accessions lists for cooperating libraries at their request.

The unique feature of this file is the inclusion of cataloging rules and guides to aid the cataloger in preparation of material for the automated catalog.

[1] Since it is not clear to what extent the automated files would include the National Union Catalog, the authority file described in this section is based on the Library of Congress collection. No attempt has been made to estimate the size of an authority file for the entire National Union Catalog.

For example, it will probably be necessary in the automated catalog to code the information inputs with special symbols, e.g., for author, editor, pagination, edition, and other data fields. This file would provide a list of these symbols and their referents, as well as a complete list of the rules for cataloging. These rules would be displayed to catalogers in the form of a tutorial display of the proper format and symbols to aid in preparing final entries. This particular subfunction would be available to Library staff only and would be closely tied to the input editing Function J.

*Function D.* Catalog Use.

*Service provided:* Permits automated search of Library of Congress catalogs and the National Union Catalog (NUC). For certain classes of materials, supplies information about location in other libraries.

This function involves methods for organizing the stored information equivalent to the present Library and NUC files and provides search strategies for selecting pertinent items from the entire store.

The structure of the catalog developed for costing purposes involves a central file, in digital form, containing complete bibliographical data for each title and ordered by accession number or some scheme that is basically sequential. This central file may be approached by at least six indexes ordered by author, title, subject, LC card number, NUC location symbol, and item number with NUC location symbols. Capacity would be provided for the addition of new indexes, such as citations, references, or key title words. These new search modes would be added as the necessary information becomes available and as the need for them becomes explicit. The indexes would be ordered alphabetically or numerically and would require periodical merging.

## *Staff and Reader File Access*

*Function E.* File Access for Staff (via input/output consoles).

*Service provided:* Provides staff members whose work involves entering, changing, or removing file entries a method for gaining access to digital catalogs and files.

This console would be used when staff members needed to enter new information into the system or to query other library files. (It should be noted that staff members whose duties require access only would probably use the less expensive read-only files provided for readers.) Although this console permits input and output, the input functions would be carefully restricted to certain specified personnel who would have "keys" to the input portion of the console in order to maintain file integrity. Certain direct input modes would allow for limited correction without requiring the intermediate step of input editing by the editing function. The latter would be required for all major entries, e.g. new titles or subject headings added to the catalog. The limited correction would include such operations as indicating the receipt of material from binding, the posting of a serial issue, etc. Thus the workload for the editing and conversion processes would be minimized by providing this direct input for limited corrections, where only an identifying code need be changed and no additional "text" is involved.

The console would provide a limited tear-off hard copy for the use of staff members. It would not, however, provide the more extensive facilities required for printing bibliographies and other extended listings.

A variety of staff consoles would undoubtedly be needed to handle the diversity of input requirements. The extent to which consoles differ according to these requirements cannot be determined without further study.

*Function F.* File Access for Readers (via output console).[2]

*Service provided:* Permits read-only access to the automated catalogs and files; provides a visual display of catalog entries, and a limited "tear-off" hard-copy production of catalog entries.

This output console would be available to readers and staff members use where there is no need for input to the files. In the early stages of the system, all consoles would probably be operated by specially trained staff members for library users, although with continued experience the console should become sufficiently self-explanatory to permit its use directly by library users. One of the console design requirements is that the console be as simple in operation as possible to minimize training required by staff and readers for its efficient use.

Provision should be made to permit a hard-copy facility (perhaps as much as 50 or 100 bibliographic items) to be printed out at the console. The reader console would not provide extensive printouts for bibliographies or searches; these would be channeled through Function J.

[2] Identified as reader console in the block diagram.

## *Circulation and Control Functions*

*Function G.* Control of Library Holdings.

*Service provided:* Provides a complete and current record of the location of all items in the automated catalog. All changes in shelf location and the status of all books temporarily not-on-shelf are recorded.

The two essential elements of this function are a location directory and a not-on-shelf (NOS) record. Under the present system, a location directory is not required for the bulk of the collection, since the classification number of the item indicates its relative location in the stacks. As subject-related classification is abandoned as a location identifier, the location directory will relate the unique identification number of each item to its location in the stacks. A more flexible storage pattern will result, since it would no longer be necessary for the relative location of a specific book to be tied to the catalog record. If considerations of frequency of use indicate that a book should be relocated more or less conveniently in the stacks, this may be effected simply by changing the location number assigned to the book's identification number in the location directory. In addition, this directory will record all books placed in special reference collections and reading rooms.

The NOS file is essential in providing control over all items temporarily in use and not in their regular shelf locations. One of the advantages of the automated system is the rapid and thorough search of bibliographic entries; this must be matched by a thorough control over the physical items, or much of the advantage of better access is lost. Therefore this file must record the status of all items in use, including the following major categories:

1. Items on loan outside the Library.[3]
2. Items checked out for reader use in the Library.
3. Items in use by the Library staff.
4. Items outside the Library for binding or other physical processing.
5. Items awaiting reshelving or in transit within the Library.

This function is not directly associated with the automated catalog, since physical retrieval is not always requested when the catalog is consulted. When a specific item is requested, a retrieval signal would be generated, screened in the NOS file, and checked in the location directory. If the relevant item number is determined to be on shelf (i.e. no entry is posted in the NOS file), the location directory will cue the appropriate stack location so that the stack attendant may retrieve the item from the shelf.

Retrieval requests, with a borrower identification number forwarded to the appropriate stack location along with the request, would be typed automatically in magnetically readable ink and characters on small detachable labels or by some similar low-cost easily applied device. These labels would be affixed to a predetermined position on the book. Thereafter any change in the location or circulation status of the book while off-the-shelf would be recorded by placing the book label under a magnetic reading device capable of automatically transmitting the number of the book and the status change to the NOS file.

These magnetic reading devices would be placed at all points where a change in the status of books may occur, including the various charge desks, the Loan Division, the points at which books are returned for shelving, and various decks in the stacks (where the return of the book to the shelves is recorded and the label removed). In all but one case the location of the reading device would dictate a unique status code which would be automatically transmitted with the item identification number. In the case of the devices used at the charge desk, however, provision would be made for indicating on a dial or key basis some further codes to indicate whether the item is used by staff or readers, and if by the staff which department or division they are in.

*Function H.* Physical Storage of Items.

*Service provided:* Storage of library materials for efficient and rapid retrieval. Ultimately may involve microstorage techniques as well as automated stack retrieval.

The adoption of an automated system will require that the Library stacks be closed in order to insure the accuracy of the various recording functions. Closing the stacks will result in a reduced need for subject-related classification as a medium for stack arrangement, since the stacks will no longer serve as a single large browsing collection. As a result, new methods of efficient storage based on demand frequency or other criteria will become feasible.

Over a period of time, the records of use generated through function M, Monitoring Use of the Collection.

[3] The Library of Congress lends material outside the Library only to Members of Congress and their staffs, certain Government officials, Government agency libraries, LC staff members, and other libraries. The term borrowers used herein refers to the above categories and not to readers who are allowed to use material in the buildings only.

should develop guidelines by which convenience of location can be made a function of frequency of use. Flexibility in storage would be further facilitated by the location directory in which storage locations can be successively assigned as needs and policies change.

At this time it does not appear feasible for the Library to consider an extensive revision of the physical storage as a condition for the installation of an automated system. It is expected, however, that, in time, the greater knowledge of use patterns made possible by the automated equipment would allow the Library to reorganize the physical collection to take maximum advantage of space and to provide maximum convenience to staff and readers. No analysis of the cost of various possible automated shelving and storage devices is given in this report; however, Function H was included because it is an area where experimentation and ingenuity might produce improvements in line with the automation of other Library services.

*Function I.* Circulation Control.

*Service provided:* Control for data input on circulation and service charging.

This function would largely be an extension and elaboration of the partially automated charging and recall process now used by the Library's Loan Division. Present procedures involve keypunching charges and machine printout of recall lists for items lent to Government agencies only. This procedure would be made more automatic and extended to cover all loans.

In the automated system, loans would be recorded by borrower and by item number. The control would obviously be closely tied to Functions G and M, since each item borrowed would be an input to the NOS file (G) and the record of item use (M). Recall of overdue items would be semiautomatic, with a computer program which would print out at regular intervals, by borrower, lists of items due or overdue.

## *Editing Functions*

*Function J.* Editing for Input and Output (Publication).

*Service provided:* Automatic editing, including selection of proper formats, etc., for new input entries added to the automated catalog. Automatic selection and editing of catalog entries for publication.

This function includes all operations involved in automatic editing of publications generated from the catalog. These publications include catalogs, demand bibliographies, and special listings beyond the scope of the console printout facilities.

For publications, the editing processor would generate a machine-readable record from the catalog, order entries in the required manner, and specify the proper format and type fonts for each element of the entry. It would also verify that each entry had the requisite number of elements such as author, title, date, etc., according to the specifications for the particular publication concerned. Many of the recurring publications would have permanent editing programs; general programs would be written for demand bibliographies for specific uses.

Subsequent to the installation of the automated system, provision may be made for incorporating this function with Function K, the processing of digital text material and commentary. This would facilitate the complete automatic production of bibliographies, including annotations, introductions, etc. This might be accomplished with Flexowriters or similar devices combined with added special editing programs.

It should be made clear that this function would not embrace the actual preparation of mats or the actual printing of general or specific Library publications. This function is concerned solely with the production of copy in proper format; the actual publication or mat preparation is included in Function K. The editing processor would generate copy in machine-readable form to be further processed by an automatic photocomposition mechanism which would then produce mats for the actual publication.

The second aspect of this function involves the editing and formatting of entries for the automated catalog. This function serves as an intermediate step between the input produced by the cataloger and the formal incorporation of the digital record into the catalog. This editing function would "shred-out" the index entries required for each new unit record. The unit record would be checked to determine whether or not all the appropriate information had been included with the proper field symbols. Errors or omissions would be relayed for display at the cataloger's console.

*Function K.* Publication of Bibliographic Materials.

*Service provided:* Automatic publication of cards, bibliographies, catalogs, and indexes by the preparation of printing masters or mats from machine-readable copy.

This function is primarily concerned with the preparation of mats or printing masters from the paper tapes or other machine-readable copy produced through the editing function (J). The hardware associated with this function is an automatic photocomposition device capable of producing masters to be further processed by the Government Printing Office.

This photocomposer must have a wide variety of type fonts and sizes and be capable of generating high quality printing. Only those Library publications prepared for permanent reference use, such as catalogs and extensive bibliographies, would require high quality printing. Less permanent productions would probably get a minimal standard printout form, perhaps not even requiring processing outside the Library. This function may also be extended to include generation of copy for Library of Congress catalog cards ordered by other libraries. In this case, the card processing function would probably include a high-speed permanent program in the editing function (J).

*Function L.* Production of Photocopy.

*Service provided:* Provides hard copy of items that cannot be circulated because of their physical condition.

This function includes microfilming and photocopying Library materials that cannot be circulated or used in their original form because of bulk, fragility, or because they exist only in microreproductions.

Initially, this function would not differ significantly from the present system of hard-copy reproduction provided by the Photoduplication Service. Advances in print-reading techniques and other improvements in photocopy technology would be incorporated in the system as they become available and as demand for such service warrants.

When published material becomes generally available in digital form and is incorporated into the automated catalog (see Function U), the production of hard copy would become fully automated.

## *Administrative Functions*

*Function M.* Monitoring Use of the Collection.

*Service provided:* Provides historical data of the use of each item in the automated Library.

As items are retrieved from physical storage a use count would be generated automatically. The input to the use file could come either from the initial retrieval request or from the magnetic reading devices at charge desks.

It is probable that no attempt would be made to distinguish between the various purposes for which items would be used, except for staff versus nonstaff use. To some extent, the circulation records of books lent outside the Library would also serve to indicate the different uses and users for specific items.

This function is closely allied with the location directory and the NOS file which are associated with Function G. One of the requirements for Function G is the closing of library stacks; maintenance of item use counts is obviously dependent on a closed stack operation.

Various data could be generated as part of Function M which would be of benefit to the Library administration. For example, statistics would be available for the number of books circulated by day and/or time of day. Such data would aid in minimizing system overloads by distributing low priority requests to periods of reduced demand and by optimizing distribution of deck attendants.

*Function N.* Monitoring Catalog Use and Response.

*Service provided:* Provides data on catalog use and response for reporting, quality control, and catalog modification.

Catalog use would be monitored and data collected for certain specified categories, such as the following:

1. Use by search mode, e.g., author, author-title, subject.
2. By user category (staff or nonstaff, and, if staff, by function, e.g., cataloging, telephone reference, LRS, etc.).
3. By priority (optional). Certain uses may have priority for processing, e.g., congressional requests.
4. By output mode requested (display, hard copy, or both).
5. By catalog response, e.g., completed search, negative response, etc.

With respect to the last item, it is assumed that the automated catalog would have preset responses when search results were negative. Most, but not all, of such negative responses would be useful in measuring the adequacy of catalog performance. Such responses could be coded, stored, and printed out along with the verbatim query, instead of merely being tallied. This may indicate areas for improvement in cataloging or for additional search modes, e.g., automatic searching under variant spellings of a name,

redefinition of a subject heading, or additional refinement in instructions for catalog use.

*Function O.* Business Data Processing.

*Service provided:* Business accounting concerning acquisitions, interlibrary services, and other bibliographic services.

This function includes all bibliographic services for which the Library charges the users or pays dealers. This function will administer billing and accounting for the following categories:

1. All purchased and priced-exchange items acquired or sent out by the Library.
2. All fee services provided for individuals, other libraries, or Government agencies, including photocopy, demand bibliographies, special searches, etc.
3. Sales of LC publications not distributed by the Government Printing Office.
4. Sale for LC printed catalog cards.

This function is restricted primarily to on-line or information handling services performed by the Library. It is not within the scope of this study to consider the automation of purely administrative accounting for payrolls, equipment, funding, etc.

(NOTE: The Library of Congress is acquiring a small general-purpose computer and related data processing equipment in January 1964. This system will be used for general Library accounting including some of the data-processing activities noted above. It is possible that, by the time a general automated system is installed, most, if not all, of the operations associated with Function O may already be automated and only integration of the existing equipment with the general system will be required.)

*Function P.* Maintaining Exchange and Gift Records.

*Services provided:* Provides records of items available for exchange, exchange agreements, items sent out on exchange, items requested and received on exchange, and gift sources.

This file would embrace most of the outflow activities and some of the inflow activities associated with the exchange and gift programs of the Library. In addition to maintaining a partial list of items available for exchange,[4] this function would include the following operations:

1. Maintain records of all exchanges and gift sources with information on addresses, key personnel, areas of special interest, and type and date of exchange agreement—all with coded field symbols and appropriate code abbreviations.
2. Periodically generate lists of items forwarded to exchange partners by name of institution, with information on the date, number, and/or value of items accepted.
3. Generate lists of items received by the Library from exchange and gift sources, and generate automatic acknowledgement of such receipts.
4. Provide for periodic and automatic review of the status of exchange agreements with respect to activity and value of such agreements.

Of the four operations listed, all but the first require extensive access to the automated records associated with monograph and serial acquisition (Functions A and B). Although exchange items forwarded for incorporation in the Library collection would be entered in those acquisition files with appropriate exchange codes, the administrative control of the exchange function would be primarily centered around files associated with Function P.

*Function Q.* Scheduling Library Operations.

*Service provided:* Accounting, status reporting, and scheduling of Library operations.

This function is essentially one of management controls. All sequential processes would be analyzed periodically in order to inform each succeeding stage of any workload variations. Trends in Library use would be analyzed to optimize space and labor allocations. Special programs would analyze data forwarded from various operating divisions of the Library through monitoring of the catalog and consoles. This function could be expanded as finer management tools were required to provide an effective liaison between the automated system and the Library administration.

## *Miscellaneous Functions*

*Function R.* Providing Special Graphic Retrieval.

*Service provided:* Provides access to a limited collection of documents and permits reproduction on console or by hard-copy device.

This function represents a type of service which could be provided in the automated library if demand warrants. For the purpose of this report, the application chosen for costing purposes was the file maintained by the Legislative Reference Service for ready reference in response to congressional inquiries. Other applications to be developed later include files main-

[4] Most of the exchange outflow consists of U.S. Government publications.

tained by the Defense Research Division, the Aerospace Information Division, specific Law Library collections, and other divisions concerned with servicing pictorial or subject-oriented materials.

At present, the Legislative Reference Service clipping file and the other special files noted above are maintained separately from the Main Catalog. It is expected that this separation will continue. This function would be largely independent of the system as a whole; however, when material in these separate files had relevance to a query put to the automated catalog, the user would be referred to the special files.

*Function S.* Communication with Other Institutions.

*Service provided:* Involves direct communication between the Library of Congress and cooperating libraries and Government agencies. Provides direct access to Library of Congress catalogs for output-only information.

This regional communications network would link the Library of Congress with other cooperating libraries and with special users, such as Members of Congress and Government agencies. The regional terminals would be similar to the output-only consoles used in the Library. In addition to queries directed to the central catalog, provision would be made for communication concerning interlibrary loans or general reference questions not specifically requiring use of the catalog.

While it is anticipated that the regional communications terminals eventually would connect directly with the automated catalog, in the initial stages communication between regional terminals and the catalog would undoubtedly require an intermediate step with, for example, the Library of Congress staff receiving and replying to these queries.

*Function T.* Converting Existing Files to Machine-Readable Form.

*Service provided:* Provides means for converting existing Library records (e.g., catalog cards) to machine-readable form and for incorporating prior holdings into the automated system.

If devices capable of automatically converting printed library cards to machine-readable form become available, this function will become essentially automatic, otherwise presently available nonautomatic or semiautomatic techniques will continue to be used.

For the purpose of costing, it has been assumed that the conversion process would be, for the most part, a straightforward translation of the old catalog cards. There would be no overall program of editing or revision, even though there may be certain inconsistencies in cataloging style. It is expected that the uniform editing provided in Function J, as well as certain more or less independent checks within the automated catalog, would screen most of the important errors that have accumulated in the catalog over the years.

The conversion of existing catalog files to digital, i.e. machine-readable form, might not need to begin until the system is in operation. As previously mentioned, certain files, such as the authority files, must be converted before the system can become efficient. The automated catalog at its inception could be essentially empty with no retrospective records. There would then be a simultaneous program of entering new records and converting old records at some predetermined rate. Exactly how this conversion would proceed has not been determined, pending further investigation into the use of the catalog and the collection with respect to date of imprint and other variables. Although the criteria for selection have not yet been determined it is anticipated that there would be some systematic and continuing conversion of the existing catalog, so that the automated catalog would expand both from new entries and conversion of old entries from the date of its inception.

*Function U.* Processing Materials Received in Digital Form.

*Service provided:* Processing of published material obtainable in digital form for inclusion in the automated library system.

When published material becomes available to the Library in digital form from publishers, and when the necessary techniques are developed, this material will be processed to automatically or semiautomatically provide entries for inclusion in the automated catalog. Also, as equipment capable of automatic conversion of print to machine-readable form for many different typefaces and sizes becomes operational, this equipment would be used to implement this function.

*Function V.* Controlling Intrasystem Communication.[5]

[5] This is the only function for which there is no corresponding block in the functional block diagram in Exhibit 1. In the block diagram the lines and arrows indicating the flow of information provide a better graphic analogy to this function than would an additional block.

*Service provided:* Routing of console requests to designated digital storage files and control of queuing by means of an intercommunication switching network.

This function is roughly analogous to that of an automatic telephone switchboard. In addition to routing messages between consoles and storage files, it will also regulate the flow of messages to any one file through the use of automatically emitted requests to wait (or "busy signals").

# *Statistics for Key Library Operations*


## *Summary*

An analysis of key library operations was made in order to project the probable throughput and data-processing workloads for the automated system. These projections indicate that, in 1972, the Library will catalog 120,000 new titles, receive and search more than 2 million serial items, and distribute more than 51 million catalog cards. In the same year it is estimated that a total of almost 4 million catalog uses will be made by readers and reference librarians. A discussion of the available data and the method of arriving at these estimates is given where necessary.


## *Introduction*

Statistics of Library operations appear in the *Annual Report of the Librarian of Congress* and in the annual reports of various divisions and departments. These statistics necessarily reflect present administrative structures and work units, with the result that much of the data is not applicable to the analysis of the requirements of the automated system. Adjustments were made when present operations would be eliminated or greatly modified by automation or when the operation includes, but is not identical with, operational units relevant to the automated system. Examples of these data include processing statistics which include operations for items, e.g. Oriental materials, not expected to be included in the automated library in its early stages.

Statistics are presented only for those operations which are relevant either to the storage capacity required in the digital files or to the frequency of access required in the automated system. Most of the projections developed are based on linear regression analyses; in general, only those projections with a correlation coefficient of .9 or greater were used. In instances where there was little correlation in the time series data, the subjective judgments of Library personnel were relied on, or estimates were developed on other bases; in either case, such estimates are clearly labeled in the exhibits.

## *Acquisitions and Internal Processing Activities*

The automated storage files associated with the acquisition and recording functions would record the recommendations for new acquisitions; the transfer of acquisitions to the cataloging divisions, bindery, or other appropriate location; the items recataloged or edited; and the receipt and holdings of serial titles. Statistics for these activities are given in Exhibits 8 and 9. In cases where the regression analysis did not indicate a strong correlation, e.g. the $r=.72$ for items recommended, the predicted values given in Exhibit 8 (and in other exhibits) were used only after approval of the top administration of the pertinent divisions.

The number of inquiries answered in connection with processing activities have not behaved in a regular pattern; between 1956 and 1960 the number was fairly constant, but there were sharp increases in 1955 and 1961. In the absence of a meaningful trend, and because of the limited range of variation, a conservative estimate of 20,000 inquiries in 1972 (the annual average for 1955–61) is used.

Similarly, the data for titles recataloged and edited show no systematic variation and again an arbitrary estimate of growth was made. In this case, an increase of 50 percent is projected, resulting in an estimate of 31,000 titles recataloged and edited in the automated system.

Over the years, the number of volumes bound seems to have increased quite steadily. Therefore, the projection of 150,000 volumes for 1972, based on the correlation coefficient for the time series regression of .88 is deemed adequate.

EXHIBIT 8.—STATISTICS FOR ACQUISITION AND PROCESSING ACTIVITIES: 1949–61

| Year | Recommendations [1] (000) | Inquiries answered (000) | Titles recataloged and cooperative editing (000) | Volumes bound (000) | New titles cataloged [2] (000) |
|---|---|---|---|---|---|
| 1961 | 123.5 | 27.98 | 20.4 | 112.5 | 90.4 |
| 1960 | 100.9 | 24.32 | 21.1 | 111.9 | 87.9 |
| 1959 | 71.2 | 22.82 | 24.8 | 90.4 | 94.0 |
| 1958 | 73.2 | 24.74 | 21.6 | 82.3 | 81.2 |
| 1957 | 70.2 | 24.08 | 18.3 | 70.5 | 72.0 |
| 1956 | 66.9 | 24.21 | 20.3 | 68.5 | 72.6 |
| 1955 | 64.6 | 24.77 | 20.3 | 61.4 | 76.9 |
| 1954 | 33.4 | 21.84 | 20.3 | 58.5 | 71.3 |
| 1953 | ..... | ..... | 19.3 | 61.2 | 68.0 |
| 1952 | ..... | ..... | 20.1 | 62.4 | 73.2 |
| 1951 | ..... | ..... | 16.7 | 56.8 | 74.7 |
| 1950 | ..... | ..... | 15.2 | 64.9 | 70.5 |
| 1949 | ..... | ..... | 23.4 | 52.3 | 58.9 |
| Correlation coefficient | r=.72 | ..... | ..... | r=.88 | r=.86 |
| Predicted value: 1972 [3] | 210,000 | 20,000 | 31,000 | 150,000 | [4] 120,000 |

[1] Includes recommendations received from the Reference Department and the Law Library.
[2] Descriptive Cataloging Division.
[3] Based on linear regression projection associated with the given correlation coefficient. Predicted values for columns two and three are based on estimates as explained in text.
[4] Predicted value of 126,000 was adjusted to 120,000 after discussions with Descriptive Cataloging Division administrators.

EXHIBIT 9.—SERIAL RECORD PROCESSING STATISTICS: 1953–61

| Year | Incoming pieces (000) | New entries (000) | Inquiries answered (000) |
|---|---|---|---|
| 1961 | 1,774.8 | 13.6 | 52.2 |
| 1960 | 1,830.3 | 12.5 | 52.1 |
| 1959 | 1,660.3 | 12.7 | 43.3 |
| 1958 | 1,472.5 | 14.5 | 42.0 |
| 1957 | 1,631.0 | 19.9 | 44.7 |
| 1956 | 1,791.2 | 16.4 | 48.7 |
| 1955 | 1,668.7 | 18.2 | 45.9 |
| 1954 | 1,505.6 | 15.5 | 38.7 |
| 1953 | 1,318.0 | 12.6 | 33.7 |
| Correlation coefficient | r=.67 | r=.31 | r=.83 |
| Predicted value: 1972 | [1] 2,100,000 | 20,000 | 70,000 |

[1] Predicted value of 2,244,000 was adjusted to 2,100,000 on basis of discussions with Serial Record Division administrators.

A similar linear regression with a correlation coefficient of .86 provides a basis for projecting statistics on the number of titles cataloged. The resulting estimate of 126,000, which would represent more than a 25 percent increase in cataloging output by 1972, was adjusted to 120,000 after discussions with Descriptive Cataloging Division administrators. Among the reasons for this adjustment was the lack of space to accommodate the number of staff members which such a growth would imply.

A significant portion of the acquisition and recording function of the automated library would be devoted to the processing of incoming serial pieces. Exhibit 9 presents a projection of the data for the number of serial pieces received, new entries added to the serial record, and inquiries answered concerning serial holdings. There has been a general, but not consistent, upward trend in the number of serial pieces received. The relevant linear regression line has a corresponding correlation coefficient of only .67. In view of this relatively low correlation, an estimate of 2,100,000 incoming pieces for 1972 was agreed upon by Serial Record Division administrators as reasonable.

The number of new entries added to the serial record each year has not behaved regularly in the period for which statistics are given. The resulting 1972 projection, which cannot be justified by the low correlation, was nevertheless accepted as a reasonable estimate. The number of inquiries answered by the Serial Record Division has behaved more regularly over the past 8 years and, as a consequence, the resulting projection of 70,000 inquiries may be considered reliable.

## *Cataloging Activities*

The analysis of cataloging and catalog activities provides estimates for several functions of the automated system including Function D, Catalog Use, and Function J, Editing for Input and Output. The data pertinent to cataloging in Exhibit 8 are used to estimate hardware requirements for these functions of the automated system.

The editing of new titles for the National Union Catalog is somewhat analogous to the cataloging of new acquisitions. It has been assumed that in the automated Library all National Union Catalog (NUC) reports of new titles would be edited and entered directly into the automated catalog. Under the present procedure there are several transfer operations before the card reaches the file from which the published catalog is generated, since the present operation has two separate routines according to the date of the publication reported. With automation this procedure would be greatly simplified and a program could be devised to do the sorting by date. The merging of the Library of Congress catalog and the National Union Catalog would obviate the need for processing new LC titles for inclusion in the National Union Catalog.

Since the present NUC statistics do not differentiate between titles reported by the Library of Congress and those reported by other libraries, it is difficult to ascertain the net new non-LC titles added each year. After an analysis of the available overlapping statistics and discussions with administrators of the Union Catalog Division, an estimate for 1961 of 87,000 new, non-LC titles was developed. The projection of this estimate for 1972 was accomplished by assuming that the same proportional increase would apply to new NUC titles as was projected for the Library of Congress. The estimate of 118,000 new, non-LC titles for 1972 was confirmed by the division.

Before new titles can be identified as such, they must be searched in the National Union Catalog and subsidiary files. All incoming reports are thus searched and confirmed either as new titles or as added locations for old titles (in the latter case they are posted on the existing card using an established location code of alphabetic characters). The number of incoming cards has increased fairly steadily over the last 10 years, as indicated in Exhibit 10. For the period 1952–61 there is a correlation coefficient of .98; the resulting linear projection for 1972 is 2,042,000. The Division administration taking into account the fact that new libraries will be reporting, as well as other factors, suggested that this figure be raised to 2,400,000.

EXHIBIT 10.—NATIONAL UNION CATALOG STATISTICS: CARDS SUBMITTED BY OTHER LIBRARIES: 1949–61

| Year | Number of cards submitted by other libraries (000) |
|---|---|
| 1961 | 1,078 |
| 1960 | 1,048 |
| 1959 | 1,077 |
| 1958 | 919 |
| 1957 | 775 |
| 1956 | 664 |
| 1955 | 549 |
| 1954 | 550 |
| 1953 | 503 |
| 1952 | 475 |
| 1951 | 830 |
| 1950 | 1,201 |
| 1949 | 867 |
| Correlation coefficient | r= .98 (1952–61) |
| Predicted value: 1972 | [1] 2,400,000 |

[1] Predicted value of 2,042,000 was revised to 2,400,000 at the suggestion of Union Catalog Division administrators.

## *Catalog Use*

The expected frequency of automated catalog use is of particular importance in estimating the hardware requirements for, and the cost of, the automated system. Use of the catalog in connection with processing and cataloging activities has been presented in preceding exhibits; catalog use in connection with other staff activities is analyzed in this section.

The use of the catalog in connection with various reference activities is presented in Exhibit 11. Note that these figures do not reflect total reference services in the three categories tabulated, but only those requests estimated to have required use of the catalog. The increase in telephone requests has been quite regular from 1949 through 1961 as indicated by a correlation coefficient of .91. The linear projected volume of 74,900 telephone inquiries and the 187,000 related catalog uses was considered sufficiently accurate. The annual volume of correspondence declined somewhat from 1949 to 1953 and then began a steady and regular rise. After 1953 the correlation coefficient is .98, the figure used to estimate the total use based on the linear projection. The data which reflect reference

help given to readers in the Library (in-person reference) have behaved quite irregularly, due partly to administrative changes, e.g. changes in hours of service, exclusion of high school students, etc. Consequently data for in-person reference service were treated parametrically, as was the reader use of the catalog.

In the absence of relevant statistics, an estimate of the number of items retrieved from the stacks for all purposes was determined indirectly. Retrieval for use within the building was derived by adding the total number of items delivered by Stack and Reader Division personnel to staff and readers and an estimate of the items retrieved by staff members themselves, and subtracting from this the number of items delivered to the Loan Division for use outside the building. A rough estimate made by the Office of the Information Systems Specialist determined that ¾ of the items retrieved for in-building use are for readers and the remaining ¼ are for staff use.

It was further assumed that, on the average, readers use the catalog four times for each book delivered.[6] Based on this assumption, the 788,200 books delivered to readers required almost 3.2 million uses of the catalog. Since no trend in the number of volumes delivered by the Stack and Reader Division is discernible, projections to 1972 were not made.

Although the automation of the Card Division is beyond the scope of this report, it is a distinct possibility, and, depending upon the system chosen, provisions for querying the central catalog by Card Division personnel must be anticipated in the specifications of hardware for the catalog. Assuming a system where the Card Division made use of the automated catalog, then, by 1972, the projected use should approximate the number of "first cards" sold. The present system requires various search techniques depending on the amount of information supplied by the purchaser: card number, author, title, series, etc. In the automated system, once the entry is identified in the automated catalog, either by card number, author, title, etc., the card number could be addressed on the console, together with the corresponding order number, to whatever automated equipment selects or generates the requisite number of cards.

The number of first cards sold from 1956 to 1961 see Exhibit 12) has increased in an almost perfectly linear fashion. Based on a correlation coefficient of .98, the number of first cards sold in 1972 would be

EXHIBIT 11.—ESTIMATED CATALOG USES FOR REFERENCE SERVICES: 1949–61

| Year | Number of reference questions requiring use of the card catalog [1] | | |
|---|---|---|---|
| | In-person [2] (000) | Correspondence [3] (000) | Telephone [4] (000) |
| 1961 | 102.8 | 13.9 | 41.9 |
| 1960 | 196.9 | 12.3 | 44.5 |
| 1959 | 100.0 | 11.8 | 45.3 |
| 1958 | 107.9 | 11.0 | 47.9 |
| 1957 | 103.3 | 9.3 | 45.6 |
| 1956 | 79.7 | 6.2 | 37.1 |
| 1955 | 96.0 | 6.0 | 36.6 |
| 1954 | 108.1 | 4.9 | 25.0 |
| 1953 | 130.2 | 3.7 | 25.1 |
| 1952 | 132.1 | 4.3 | 25.3 |
| 1951 | ..... | 4.5 | 25.6 |
| 1950 | ..... | 6.2 | 24.2 |
| 1949 | ..... | 7.0 | 19.2 |
| Correlation coefficient | ..... | .98 (1953–61) | .91 |
| Predicted value: 1972 [5] | [8] 180,000 | 28,600 | 74,900 |
| Weighting factor [6] | 2.0 | 5.0 | 2.5 |
| Total catalog uses: 1972 [7] | [8] 360,000 | 143,000 | 187,000 |

[1] Includes General Reference and Bibliography Division, Hispanic Foundation, Rare Book Division, Science and Technology Division, and Slavic and Central European Division.

[2] An estimated 70 percent of all in-person reference questions require use of the card catalog, based on interviews with reference librarians. This represents use of the card catalog by reference librarians on behalf of the patron, not use by the patron himself.

[3] Most letters drafted in response to reference questions require use of the catalog; form letters do not usually require use of the catalog.

[4] Interviews with telephone reference librarians indicated that approximately 80 percent of telephone reference questions require the use of some catalog (e.g. the L. C. printed catalog); only ⅔ of these, however, require the use of the card catalog.

[5] Based on linear regression analyses.

[6] The average number of individual catalog uses associated with each request.

[7] Predicted number of requests in 1972 times the weighting factor rounded to the nearest thousand.

[8] See Exhibit 15 and accompanying Note 8.

[6] This is an area where lack of detailed data made arbitrary assumptions necessary. The best available measure of reader use of the Library collection is the Stack and Reader Division count of volumes delivered; however, in order to estimate reader use of the catalog it would be better to have a figure for the number of *titles* delivered. We also know that readers use the catalog for books requested but not delivered for various reasons, and they use the catalog when retrieval of material is not desired. It was arbitrarily decided that all these uses could be approximated with sufficient accuracy by allowing 4 catalog uses per volume delivered, since that figure should allow for sufficient overestimation to account for uses not measurable by any available data. (For example, if the reader submits a slip for a 5-volume work on the history of Rome, or for 3 volumes of a particular serial, presumably he used the catalog a maximum of 4 times before selecting that particular title, but not 4 times for each volume.)

slightly more than 11.4 million. To better reflect the actual number of projected catalog uses, this estimate was adjusted to exclude cards for materials not expected to be included in the initial automated system, e.g. Orientalia, music, etc. These "excluded" cards have averaged 12 percent of the total volume of first cards since 1955. Reducing by 12 percent, the projected volume for 1972 yields an estimate of 10 million catalog uses.

EXHIBIT 12.—CARD DIVISION STATISTICS: 1949–61

| Year | Total cards distributed (000) | First cards (000) |
|---|---|---|
| 1961 | 40, 282 | 8, 001 |
| 1960 | 37, 709 | 7, 492 |
| 1959 | 35, 220 | 6, 725 |
| 1958 | 33, 322 | 7, 098 |
| 1957 | 31, 745 | 6, 359 |
| 1956 | 29, 115 | 6, 303 |
| 1955 | 27, 598 | ..... |
| 1954 | 26, 204 | ..... |
| 1953 | 25, 380 | ..... |
| 1952 | 25, 767 | ..... |
| 1951 | 24, 966 | ..... |
| 1950 | 24, 984 | ..... |
| 1949 | 25, 642 | ..... |
| Correlation coefficient | . 93 | . 92 |
| Predicted value 1972 | 51, 397, 000 | [1] 10, 067, 000 |

[1] This predicted value was adjusted from 11,440,000 to exclude retrieval of cards not expected to be included in the automated catalog, e.g. map cards, oriental-language cards, etc.

## *Circulation and Retrieval Activities*

The key statistics associated with circulation and retrieval operations are the number of volumes retrieved by the Stack and Reader Division and the additional number of items retrieved by the Photoduplication Service. Each of these is more or less directly applicable to the determination of the hardware requirements for Function G of the automated system, Control of Item Handling and Location. In the automated system, this function would also include the assignment of storage locations for new volumes added to the collection, for this reason the number of new volumes shelflisted is a pertinent figure.

All data mentioned above are shown in Exhibit 13. Since neither the items retrieved by Photoduplication nor the volumes delivered by Stack and Reader show any systematic change over time, the total number of volumes delivered in 1972 is assumed to be equal to the present total of approximately 1.1 million.

Time series data on the number of loans indicate irregular fluctuations from year to year, although a distinct upward trend is discernible. The linear projection of 238,000 for 1972 is somewhat questionable, since the coefficient of correlation is only .67, and the level of operation has actually declined since 1959. An estimate of 220,000 loans, which allows for a smaller increase was therefore used.

The shelflisting statistics (Exhibit 13) are similarly inconclusive and do not permit any meaningful projection. Elsewhere in this appendix, greatly increased workloads are projected for the closely related operations of cataloging and serial processing. Since bound serial issues and new monographic acquisitions are the primary inputs for shelflisting, an estimate of 180,000 volumes for 1972 was made to allow for this increased effort.

EXHIBIT 13.—STATISTICS FOR VOLUMES DELIVERED, RETRIEVED BY PHOTODUPLICATION, SHELFLISTED, AND LENT: 1949–61

| Year | Items retrieved by Photoduplication Service (000) | Volumes delivered by Stack and Reader Division (000) | Volumes shelflisted (000) | Loans (000) |
|---|---|---|---|---|
| 1961 | 140. 7 | 955. 4 | 106. 0 | 202. 1 |
| 1960 | 110. 3 | 956. 7 | 107. 2 | 202. 5 |
| 1959 | 95. 9 | 946. 9 | 116. 1 | 217. 9 |
| 1958 | 79. 2 | 1, 043. 1 | 104. 7 | 207. 1 |
| 1957 | 82. 5 | 945. 1 | 99. 7 | 200. 9 |
| 1956 | 87. 0 | 969. 0 | 110. 7 | 205. 5 |
| 1955 | 93. 8 | 957. 6 | 110. 1 | 183. 3 |
| 1954 | 148. 2 | 1, 016. 9 | 96. 8 | 201. 8 |
| 1953 | 84. 0 | 1, 032. 8 | 108. 9 | 196. 2 |
| 1952 | 68. 1 | 1, 046. 4 | 110. 2 | 211. 9 |
| 1951 | 68. 6 | 1, 035. 8 | 102. 2 | 183. 6 |
| 1950 | 68. 2 | 1, 142. 5 | 103. 6 | 183. 7 |
| 1949 | 114. 6 | 1 087. 8 | 98. 1 | 167. 5 |
| Predicted values: [1] 1972 | 140, 700 | 955, 400 | 180, 000 | 220, 000 |

[1] See accompanying text for basis of predictions.

SECTION V

# *Basic Hardware Specifications*

## *Summary*

In this section the functions and workload statistics developed in preceding sections provide the basis for an analysis of the gross specifications for equipment capable of serving the automated Library adequately. This equipment includes a processing file with a capacity of approximately $9 \times 10^8$ bits with 22 uses per minute and a central catalog group with a minimum capacity of $4.7 \times 10^{10}$ bits and a proposed $10^{12}$ bit capacity (to allow for growth of Library catalog and development of finer indexing) with frequency of use ranging from 134 to 206 uses per minute. The other major hardware groups include a circulation and retrieval complex with a $1.2 \times 10^9$ bit file capacity with 39 uses per minute, 50 off-line electric typewriters, and 70 limited-capacity magnetic-print readers. The consoles proposed for the Library have a 1,000 character screen capacity with associated internal buffering; from 148 to 163 such consoles are required for basic Library activities, exclusive of fee services and network communications.

## *Introduction*

Although many general requirements for the hardware equipment were considered during this study and, to some extent, treated in this report, the primary focus has been on detailed analyses of the requirements for digital storage capacity and frequency of use (file access). The significance of these two factors is due to the large volume of bibliographical information that must be stored digitally and to the necessity for constant and rapid access to this store of information. On the basis of these requirements and the operational requirements developed in Section III, it is possible to determine generally, and in some cases specifically, the capabilities required of equipment needed for an automated Library system.

In this report "capacity" refers to fast-access storage capacity in separate digital-form files that may be associated with general computers, not to the core memory capacity of those computers themselves, and it is presented in terms of both characters and bits (in the summary exhibits only numbers of bits are used). The term "use" in the phrase "frequency of use" refers to a completed single look-up on the part of the operator or reader. This look-up may require one or several "access" cycles by the computer in the relevant storage file. Only one cycle would be required when the desired item is uniquely identified in the search request, but in many cases the desired item, or items, may be identified only by the fact that they satisfy several different search criteria. In these latter cases, a single use may require numerous cycles of the computer in order to compare several different items in different parts of the storage file.

## *The Central Catalog*

The largest single storage file in the automated system is the central catalog. This, at the maximum, could be a single integrated file containing the full contents of the National Union Catalog and therefore, by definition, the Library of Congress main catalog. The organization and suggested search modes for this catalog are discussed in Section III of this Appendix.

At the maximum, i.e. including the National Union Catalog, this file is estimated to require a capacity of $4.5 \times 10^{10}$ bits by 1972 (see Exhibit 14). This capacity relates only to the initial size without allowance for future growth or for the addition of new indexes. It is obvious that an automated catalog would provide opportunity for the development of new indexes, refinements in cataloging, and the possible introduction of text material in digital form into the central catalog. For this reason, there is little basis for attempting even an approximate estimate of the growth of

this catalog file subsequent to its installation; obviously the hardware must be modular in design.

The cost of storage files does not vary continuously with respect to capacity. For the most part, increases in file capacity are such that, for the purpose of price comparison, a significantly large file will have a capacity 100 or 1,000 times larger than the next smaller category. At the present time, existing commercially available rapid access files do not much exceed $10^{10}$ bits of storage capacity. The next major development under consideration is a trillion-bit ($10^{12}$) read-only file. With such a file, the need for precision in the measurement of the central catalog is vastly reduced, since the trillion-bit memory would permit a twenty-fold expansion in the initial size of the central catalog. (These conclusions also apply to the other storage files discussed in this section. Possible errors in the capacity estimates are not particularly significant at the margin and will affect costs only if there are errors inducing order-of-magnitude changes.)

Exhibits 15 and 16 show the frequency with which the central catalog is expected to be used. The total annual use estimated for different categories reflects differing total hours of operation, e.g. some units of the Library are open on weekends, others maintain a 40-hour workweek. Roughly estimated peak load factors were applied to the resulting frequencies per hour or per minute to account for fluctuations in the level of use. No peak loading was allowed for staff or fee service use, since it was assumed that these uses can be more easily predicted and scheduled than can those which depend on the number of readers in the Library. In addition, any special pressure on these activities could be compensated for by transferring some operations to off hours.

EXHIBIT 14.—CENTRAL CATALOG SPECIFICATIONS: INITIAL SIZE

A. *Title Count*

| | |
|---|---|
| Number of titles in National Union Catalog in 1961 [1] | 10, 500, 000 |
| Number added by 1972: [2] | |
| To Library of Congress Catalog | 1, 100, 000 |
| To National Union Catalog (other than Library of Congress) | 1, 250, 000 |
| Total number of titles in 1972 | 12, 850, 000 |

B. *Character and Bit Counts*

| File or index | Unit | Characters per unit | Number of units | Total characters |
|---|---|---|---|---|
| 1. Basic card file | Library card | [3] 275 | 12, 850, 000 | 3, 533, 750, 000 |
| 2. Author index | Author name | [3] 18 | [4] 5, 000, 000 | 90, 000, 000 |
| 3. Subject index | Subject heading | [3] 23 | [5] 285, 000 | 6, 555, 000 |
| 4. Title index | Title | [3] 40 | 12, 850, 000 | 514, 000, 000 |
| 5. NUC index by item | Location code | 6 | [6] 192, 750, 000 | 1, 156, 500, 000 |
| 6. Card number index | Card number | 7 | 12, 850, 000 | 89, 950, 000 |
| 7. Item numbers for indexes 2 to 6 | Item number | 8 | [7] 64, 250, 000 | 514, 000, 000 |
| 8. NUC index | Location code | 6 | [8] 1, 600 | 9, 600 |
| By location | Item number | 8 | 192, 750, 000 | 1, 542, 000, 000 |
| Total number of characters in initial catalog | | | | 7, 446, 764, 600 |
| Total number of bits in catalog at 6 bits/character | | | | 4.5 x $10^{10}$ |

[1] Douglas Bryant, Chairman of the Committee on the Preservation of Research Library Materials, cited in the *Library of Congress Information Bulletin*, June 18, 1962, p. 307.

[2] Increase 1961–72 estimated in the following manner:

| | Annual growth | | | 11-year increase |
|---|---|---|---|---|
| | 1961 | 1972 | 61–72 average | |
| LC Catalog | 80, 000 | 120, 000 | 100, 000 | 1, 100, 000 |
| Other NUC Catalog | 91, 000 | 118, 000 | 104, 500 | 1, 250, 000 |

[3] A study conducted by the Office of the Information Systems Specialist of the Library of Congress in 1961 indicated a mean number of characters per card of 237. This included the call number but not the spaces between words. The additional 37 characters allow for spaces and field symbols. The same study developed mean number of characters for author, title, and subject heading.

[4] Estimate by Planning Research Corporation based on twice the number presently estimated for Library of Congress Main Catalog.

[5] Estimate from actual count in 1953 plus observed and expected increases (see Exhibit 23).

[6] For each title an average of 15 locations estimated (generously) by Library personnel: 15 x 12,850,000 = 192,750,000.

[7] The character counts for items 2 and 6 reflect only the indexing descriptors and not the item numbers indexed. These item numbers are reflected as a single calculation, the number of titles times the 5 indexes per title: 5 x 12,850,000 = 64,250,000.

[8] Estimate of number of cooperating libraries by Library personnel.

Available statistics on the number of readers in the Library do not show a significant trend (Section IV). For this reason, reader use and in-person reference use of the Library are treated parametrically, as is communication with the catalog by other libraries through the regional communications network. For the reader and in-person reference use, two levels of operation were chosen for the purpose of making cost estimates: (1) the present use level, and (2) an increase of 50 percent over the present use level.

Since no historical data exist on the use of a nationwide library communication system, such as the one proposed, estimates of use were based on two arbitrary assumptions. The lower use level is based on the presence of 100 consoles at research libraries throughout the country, with each console handling 15 requests per hour. The higher level projects 200 consoles with a maximum average frequency of 20 uses per hour per console. The resulting total catalog use from all categories within and outside the Library ranges between 2.05 and 3.2 uses per second.

## *The Monograph Processing File*

This file is required for Function A, Recording of Monograph Acquisition and Internal Processing. Unlike the central catalog this file does not need to be cumulative. It need only record those items currently recommended for acquisition or items being processed prior to formal incorporation in the central catalog

EXHIBIT 15.—CENTRAL CATALOG SPECIFICATIONS: NUMBER OF USES PER YEAR

| *Type of use* | *Estimated uses per year: 1972* [1] |
|---|---|
| A. Staff Use (not including in-person reference): | |
| Cataloging | [2] 120,000 |
| Recataloging and cooperative editing | 31,000 |
| Editing new NUC titles | 118,000 |
| NUC reports | 2,400,000 |
| Reference (correspondence) | 143,000 |
| Bibliography preparation | [3] 295,000 |
| Acquisition searching | |
| Reference Department | [4] 282,000 |
| Law Library | [5] 146,000 |
| Selection Officer | [6] 50,000 |
| Decimal Classification Office | [7] 40,000 |
| Exchange and Gift Division | 100,000 |
| Legislative Reference Service | [8] 350,000 |
| Aerospace Information Division | [8] 16,000 |
| Defense Research Division | [8] 20,000 |
| Total | 4,111,000 |
| B. Reader and In-Person Reference Use: | |
| Telephone Reference use | 187,000 |
| In-person reference use | [9] 360,000 |
| Reader use | 3,152,000 |
| Total | 3,699,000 |
| C. Fee service use: | |
| Photoduplication | 394,000 |
| Card Division | 10,067,000 |
| Total | 10,461,000 |

[1] Except where otherwise noted these estimates are taken directly from the discussion of key Library statistics in Section IV.

[2] No weighting factor is used for descriptive and subject cataloging use of central catalog, since most of the use of this will be associated with the automated catalog authority file (see Function C).

[3] The number of bibliographic entries varies widely from year to year. The number of catalog uses per entry also differs for bibliographies in different divisions. Questioning of bibliographers reveals an average of 10 uses per entry in the Rare Book and Slavic Divisions, 20 in the General Reference and Bibliography Division and only 2 per entry for bibliographies prepared by the Hispanic Foundation. Since the total number of catalog uses for preparation of bibliographies indicates no regular trend over time the estimate used here (295,000 uses) is based on the highest observed figure in the period 1955–61.

[4] This estimate is based on twice the number of title searches made in 1961 (93,651). Division personnel estimate 1.5 catalog uses per title searched. The doubling of present operations is based in large part on the 70 percent increase projected for titles recommended (see Exhibit 8).

[5] This estimate was derived in the same manner used for acquisition searching in Reference Department discussed above in note (3), except for two numerical differences: Law Library personnel estimate 2.0 uses per title; number of titles searched in 1961 was only 36,657.

[6] An arbitrary assumption based on the limitation of this operation of the Selection Officer and one assistant for whom catalog searching is only a limited part of their activities.

[7] Based on an average of approximately 30,000 titles classified for the past 15 years with provision for a maximum increase of 10,000 by 1972.

[8] No direct estimates of the number of catalog uses were available. Current estimates were calculated from the total time spent in catalog searching as reported by division personnel. No projections were made since there is no systematic relationship between the use of the catalog and the primary operations of the division.

[9] Estimated at 1961 levels, allowing two catalog uses per item requested and searched in accordance with division estimates, plus additional provision for a maximum increase of 30,000 items by 1972.

EXHIBIT 16.—CENTRAL CATALOG SPECIFICATIONS: NUMBER OF USES PER HOUR AND PER SECOND

| Type of use [1] | Number of uses per year [1] | Peak load multiple factor | Hours of operation per year | Number of uses per hour |
|---|---|---|---|---|
| Staff use | 4, 111, 000 | [2] 1.0 | [5] 3, 000 | 1, 370 |
| Telephone reference use | 187, 000 | [3] 1.5 | [5] 3, 000 | 93 |
| Reader and in-person reference use: | | | | |
| Present level | 3, 512, 000 | [4] 2.5 | [6] 3, 700 | 2, 373 |
| 50 percent increase | 5, 268, 000 | ..... | ..... | 3, 559 |
| Fee service use | 10, 461, 000 | [2] 1.0 | [7] 4, 000 | 2, 615 |
| Regional network use: | | | | |
| Low projection [8] | ..... | ..... | ..... | 1, 000 |
| High projection [9] | ..... | ..... | ..... | 4, 000 |

*Total Catalog Uses*

| Assumptions | Uses per hour | Uses per second |
|---|---|---|
| Present reader use, Low regional use | 7, 451 | 2.1 |
| Present reader use, High regional use | 10, 451 | 2.9 |
| Reader use up 50 percent, Low regional use | 8, 637 | 2.4 |
| Reader use up 50 percent, High regional use | 11, 637 | 3.2 |

[1] See Exhibit 15.
[2] No peak loading provided for since it is assumed that these operations will be scheduled to minimize peaking.
[3] Estimate by Planning Research Corporation.
[4] Estimate based on hourly peak load factor of 2.46 for the number of persons using the catalog, according to a study conducted in 1961 by the Library's Information Systems Specialist.
[5] Assumes operation 12 hours per day (1½ shifts), 250 days per year.
[6] Assumes operation 73 hours per week (the present hours of the Library) for 51 weeks per year (one week deducted for holiday closings).
[7] Assumes operation 16 hours per day (2 shifts), 250 days per year.
[8] Low projection assumes 100 regional terminals with a maximum of 15 requests per hour from each.
[9] High projection assumes 200 terminals with a maximum request frequency of 20 per hour from each.

and location in the stacks. As indicated in Exhibit 17, the initial size of this file is small (3.6 million characters), since it would begin operation with only records of standing orders and of vendors used by the Library.

The annual volume of transactions indicates that this file requires a capacity of approximately 46.6 million characters, if the average elapsed time between recommendation and formal inclusion in a catalog approximates one year. Although precise data are not available on the distribution of elapsed time, Library personnel have indicated that for most current acquisitions 6 months should be a reasonable maximum for an item to be fully processed under the present system. On this basis, a 6-month turnover period in the automated system, which will be faster than the present system, should be a reasonable estimate.

Based on this turnover rate, the required capacity for the monograph processing file need only accommodate one-half the annual volume plus the initial order and vendor files, approximately $2.7 \times 10^7$ characters, or $1.6 \times 10^8$ bits. This capacity is less than 1 percent of the size of the central catalog and is well within the spectrum of presently available equipment.

## *The Serial File*

Estimates of the required capacity of the file associated with Function B, Recording Serial Acquisitions, Processing, and Holdings, are presented in Exhibit 18. One portion of this file should be a more or less permanent, cumulative record of serial titles with pertinent identifying information. That portion of the file required for posting receipt of incoming serial pieces and transmittals to the bindery is essentially a transactions record and may be considered to be of relatively constant size.

715-690—64——5

EXHIBIT 17.—MONOGRAPH PROCESSING FILE SPECIFICATIONS

| | | |
|---|---|---|
| A. *Initial size* | | |
| Standing order file: [1] | | |
| 4,000 entries at 100 characters each | | 400,000 |
| Vendor file: [2] | | |
| 16,000 vendors at 200 characters each | | 3,200,000 |
| | | 3,600,000 |
| B. *Capacity (annual volume)* | | |
| Items recommended (1972) [2] | 210,000 | |
| Characters per item recommended: | | |
| Main entry [3] | 18 | |
| Title [3] | 40 | |
| Imprint [4] | 40 | |
| Volumes | 1 | |
| Order number | 6 | |
| Status code | 2 | |
| Other information [5] | 43 | |
| Total | 150 | |
| Total characters for items recommended | | 31,500,000 |
| Title to order number file | | 9,660,000 |
| Order number to main entry file | | 5,460,000 |
| Total | | 46,620,000 |
| C. *Total required capacity* | | |
| Initial size | | 3,600,000 |
| Annual volume reduced by 50 percent assuming maximum average elapsed time (recommendation to final cataloging) is 6 months | | 23,310,000 |
| Total (characters) | | 26,710,000 |
| Required capacity (in bits) | | $1.6 \times 10^8$ |

D. *Use*

| *Category* | *Number of uses per year: 1972* |
|---|---|
| Entering recommendations | 210,000 |
| Posting order numbers or purging for all recommendations | 210,000 |
| Accessioning pieces received [6] | 294,000 |
| Posting transfer to cataloging [7] | 120,000 |
| Reference requests [8] | 35,000 |
| Posting transfer to bindery [9] | 30,000 |
| Posting receipt from bindery [9] | 30,000 |
| Recataloging and cooperative cataloging [10] | 20,000 |
| Total | 949,000 |

| | | |
|---|---|---|
| Number of hours of operation per year | | [11] 3,000 |
| Number of uses: | | |
| Per hour | 316.3 | |
| Per minute | 5.3 | |

[1] Estimates by Order Division personnel.

[2] See Exhibit 8.

[3] See character count estimates, Exhibit 14.

[4] Estimate from Library of Congress card sample.

[5] Estimate by Planning Research Corporation.

[6] Pieces accessioned exceed recommendations due to standing orders and multiple volumes (pieces) per title. This estimate was derived by applying the 1961 ratio of pieces accessioned to titles recommended to the 1972 projection of titles recommended.

[7] See Exhibit 14.

[8] See the discussion on acquisitions and internal processing statistics in Section IV.

[9] Estimated 20 percent of projected combined total of binding and rebinding in 1972. The balance of 80 percent was estimated to be attributable to serial binding. The projection of total volume was based on linear regression analysis with a correlation of coefficient $r = .86$.

[10] Assumed constant, i.e., equal to 1949–61 average. Mean absolute deviation in that period is 1,800. Includes cooperative titles adapted.

[11] Twelve hours (1½ shifts) per day, 250 days per year.

EXHIBIT 18.—SERIAL FILE SPECIFICATIONS: CAPACITY AND USE

| | | |
|---|---|---|
| A. *Capacity* | | |
| Initial size: | | |
| Number of titles in 1961 (including approximately 150,000 inactive) [1] | | 400,000 |
| Estimated increase by 1972 [2] | | 187,000 |
| Total | | 587,000 |
| Characters per serial file entry: | | |
| Title [3] | 94 | |
| Frequency of issue [4] | 4 | |
| Number of copies | 2 | |
| Routing | 5 | |
| Identification number | 8 | |
| Total | 113 | |
| Total characters | | 66,331,000 |
| Total bits [5] | | 397,986,000 |
| | | $3.98 \times 10^8$ |
| Posting capacity: | | |
| Average number of characters posted per active title: [6] | | |
| 6 issues at 10 characters each [7]=60. 337,000 active titles at 60 characters each | | 20,220,000 |
| Binding records: | | |
| 120,000 volumes [8] bound at 15 characters each [9] | | 1,800,000 |
| Total characters | | 22,020,000 |
| Total capacity: | | |
| 88,351,000 characters. | | |
| $5.3 \times 10^8$ bits.[5] | | |

| B. *Use* — *Category* | *Number of uses per year* |
|---|---|
| Search and post receipts [10] | 2,100,000 |
| Enter new titles [11] | 20,000 |
| Answer reference inquiries [12] | 70,000 |
| Post transfer to and receipt from bindery for 120,000 volumes [8] | 240,000 |
| Search New Serial Titles reports [13] | 158,000 |
| Total number of uses per year | 2,588,000 |
| Assuming 3,000 hours of operation per year: | |
| Number of uses per hour | 862.7 |
| Number of uses per minute | 14.4 |

[1] Estimate by Office of the Information Systems Specialist confirmed by division administrators.

[2] Estimated in the following manner:

| | |
|---|---|
| Present rate of annual increase | 14,000 |
| Estimated 1972 rate of annual increase confirmed by division administration | 20,000 |
| Average (assuming linearity) over 11 years. | 17,000 |
| 11-year total=average x 11 | 187,000 |

[3] Based on random sample of 100 titles from current volume of *New Serial Titles*.

[4] Includes frequency per year (maximum 2 digits) and starting date (maximum 3 digits).

[5] Assumes 6 bits per character.

[6] Maximum number of issues per title posted before next volume is bound. Averaged for all active titles.

[7] Each posting requires date (3 digits) and routing (1 to 5) digits. Assume average is 5 characters per copy and 2 copies per issue.

[8] Estimated at 80 percent of total volumes bound and rebound projected for 1972; see note (9), Exhibit 17.

[9] Three dates at 4 characters each (dates of first and last issue bound, date sent to binder), binding code (1) and number of issues (2).

[10] Estimate confirmed by division administrators.

[11] See note (2).

[12] Estimate by Planning Research Corporation based on linear regression analysis with a correlation coefficient r=.8.

[13] Estimate by Planning Research Corporation confirmed by division administrators.

It has been assumed, in estimating the required capacity for the constant portion of the file, that, for all serial titles, an average of 6 issues are posted prior to binding (which clear the record). The required capacity for the serial file is slightly more than 88 million characters ($5.3 \times 10^8$ bits) with a use frequency of 14.4 per minute.

## *The Exchange Record*

A significant portion of the serials and monographs added to the Library are acquired through exchange agreements with U.S. Government agencies, foreign governments, universities, and other libraries. Among the major problems in administration of the exchange program, according to administrators in the Exchange and Gift Division, are accurate recording of items received from exchange partners, so that these may be acknowledged and balanced against the items offered, and keeping aware of the status of the thousands of exchange agreements. In part, these problems would be solved by the incorporation of exchange partner codes in the digital records for monographs and serial acquisitions. Other aspects of exchange administration, such as descriptive information on exchange partners and the exchange balance with each, could be handled by an automated exchange record.

The capacity and use requirements of this exchange record file are presented in Exhibit 19. The permanent portion of the file, which determines its initial size, includes information on the more than 20,000 exchange sources, with codes to permit search by special interest area, geographical location, type of exchange agreement, etc.; additional capacity is required to record the annual volume of transactions. Automation of the Library would cause drastic changes in the volume and type of transaction activities, since much of the pertinent information could be searched and selected automatically from the monograph and serial acquisitions files. A systematic projection of the transaction capacity requirements would involve both a detailed elaboration of the operations of the proposed system and a specific management study of present exchange procedures. In lieu of this effort, rough, and perhaps arbitrary, estimates of both transaction capacity requirements and frequency of use have been developed, based in part on discussions with Exchange and Gift Division staff members who have administrative responsibilities.

EXHIBIT 19.—EXCHANGE RECORD SPECIFICATIONS

| | |
|---|---|
| A. *Initial size* | |
| Number of exchange and regular gift sources [1] | 23, 000 |
| Number of characters required to describe each [2] | 600 |
| Total number of characters | 13, 800, 000 |
| B. *Annual volume* | |
| Number of titles exchanged [3] | 100, 000 |
| Number of characters per single exchange request [4] | 200 |
| Total number of characters | 20, 000, 000 |
| C. *Total required capacity* | |
| Number of characters | 33, 800, 000 |
| Number of bits | $2 \times 10^8$ |
| D. *Use* | |
| Assume 5 uses per title: [5] | |
| 500,000 uses per year. | |
| Assume 3,000 hours of operation per year: | |
| 166.7 uses per hour. | |
| 2.8 uses per minute. | |

[1] Although based on present levels of activity this estimate provides for future expansion since only about 10 percent of the total number of exchanges (22,300) are presently active.

[2] Estimate by the Office of the Information Systems Specialist.

[3] Based on an Exchange and Gift Division estimate of 50,000 titles in FY 1961, doubled to allow for future growth.

[4] Estimate by Planning Research Corporation. Author, title, identification, and exchange codes and dates of order and receipt together add to approximately 100 characters. An additional 100 are provided for other information.

[5] Estimate by Planning Research Corporation.

## *The Location Directory and the Not-on-Shelf File*

Function G, Control of Library Holdings, requires two related digital storage files: the location directory and the not-on-shelf file, for which capacity and use requirements are given in Exhibits 20 and 21, respectively. It is not anticipated that a major portion of the Library collection would need to be relocated as part of the transition to an automated system. However, the closing of the stacks, which is anticipated, would permit more efficient shelving of new acquisitions and, as item-use data are generated by the system, the gradual and systematic relocation of holdings according to criteria, such as frequency of use and size.

The annual growth of this directory file depends on the number of volumes relocated as well as the number of new volumes acquired and therefore cannot be predicted accurately at this time. The growth estimates shown in Exhibit 20 should be interpreted as a reasonable guess at the possible growth. The maximum file

capacity, generously estimated to provide for a collection twice the present size, with specific location numbers for every volume, is about 2 x $10^8$ characters or about 1.2 x $10^9$ bits. This file would probably be used slightly more than 18 times per minute.

The not-on-shelf file, closely associated with the location directory, would be consulted every time a retrieval request was made. Since it would be a current file of books not in the regularly assigned locations in the Library, it would be a small file with a more or less constant size. Estimated capacity for this file is 4.7 million characters or 2.8 x $10^7$ bits (see Exhibit 21), about 2.5 percent of the size of the location directory.

## *The Circulation File*

Records for all items lent outside the Library would be kept both by item number and by a borrower identification code or name. The computations by which the capacity and use requirements for this file were developed are shown in Exhibit 22. The total required capacity of the circulation file is 1.8 x $10^7$ bits, roughly equal to that of the not-on-shelf file. Frequency of use for this file is estimated to be 2.4 per minute. The foregoing estimates have been based on present levels of operations, and no projections have been developed for an increase in the number of loans, since the number of loans per year is largely determined by administrative factors, such as restrictions on material that may be lent and on individual or institutional borrowers.

## *Catalog Authority File*

The catalog authority file, associated with Function C, lists the author entries, subject headings, and cataloging terms in the form established by the Library of Congress as a standard for its own cataloging and widely used by libraries throughout the country. Exhibit 23 indicates the requirements for capacity and access to this file. The initial capacity is estimated from data on present size and from projected growth trends; the maximum capacity, arbitrarily set at twice the 1972 capacity, is 3.6 x $10^9$ characters, or 2 x $10^9$ bits.

EXHIBIT 20.—LOCATION DIRECTORY SPECIFICATIONS

A. *Initial size:*
Effectively zero; small directory relating existing stack location numbers to the existing Library of Congress classification schedule.

B. *Annual growth:*

| | |
|---|---|
| New volumes shelflisted [1] | 180,000 |
| Volumes relocated [2] | 100,000 |
| Total | 280,000 |

Entry size (for each volume): [2]

| | *Characters* |
|---|---|
| Item number | 8 |
| Deck number | 2 |
| Shelf number | 3 |
| Copy number | 2 |
| Total | 15 |

Total annual growth:
4,200,000 characters.
2.52 x $10^7$ bits.

C. *Estimated maximum capacity required:*

| | |
|---|---|
| Number of volumes in Library: 1961 [3] | 6,440,000 |
| Estimated maximum volumes (twice present collection) | 13,000,000 |
| Estimated maximum characters in location directory at 15 characters per volume entry | 195,000,000 |
| Estimated maximum number of bits | 1.2 x $10^9$ |

D. *Use:*

| | |
|---|---|
| Number of retrievals [4] | 1,100,000 |
| Number of not-on-shelf responses [4] to retrieval requests | 240,000 |
| Number of volumes shelflisted [1] (combined new and relocation) | 280,000 |
| Total | 1,620,000 |
| Assuming 3,700 hours of operation per year: [5] | |
| Number of uses per hour | 432.8 |
| Number per minute | 7.3 |
| Peak load factor [6] | 2.5 |
| Peak use frequency (uses per minute) | 18.3 |

[1] See the discussion of circulation and retrieval activities in Section IV.

[2] Estimate by Planning Research Corporation.

[3] *Annual Report of the Librarian of Congress* for FY 1961. Refers to classified collections only.

[4] Based on the 1961 volume for the combined total of Stack and Reader Division and Photoduplication Service which retrieves items that are to be reproduced. No projection was made, since levels vary sharply and unsystematically from year to year, and what little trend can be seen seems slightly downward (see Exhibit 13).

[5] Reflects hours Library is open (73 hours per week; 51 weeks per year). This is somewhat high for staff use (estimated at 3,000 hours per year) but is exactly equal to reader and in-person reference use.

[6] Based on peak loads in reader use (see note (4), Exhibit 16). Since a portion of total retrieval is for staff use, and hence is not subject to such large fluctuation, this estimate probably overstates use frequency and compensates for the high estimate of hours of operation (see note (6) above) which would tend to overstate use frequency.

EXHIBIT 21.—NOT-ON-SHELF FILE SPECIFICATIONS

A. *Initial size*

Transfer of "hard core," long-term intralibrary charges from Central Charge File—200,000 entries.[1]

B. *Current file*

| Type of charge | Annual volume | Minimum annual turnover [4] | Peak load factor | Maximum required capacity (entries) [7] |
|---|---|---|---|---|
| Loan | [2] 220, 000 | 10 | [5] 1. 5 | 33, 000 |
| Staff | [3] 300, 000 | 6 | [5] 1. 0 | 50, 000 |
| Reader | [3] 600, 000 | 150 | [6] 2. 5 | 10, 000 |
| Total | | | | 93, 000 |

C. *Total required capacity*

| | | |
|---|---|---|
| Entries: | | |
| Initial file | 200, 000 | |
| Current file (maximum) | 93, 000 | |
| Total | 293, 000 | |
| Characters per entry: | | |
| Item number | 8 | |
| Status (location in Library) | 4 | |
| Date charged | 4 | |
| Total | 16 | |
| Total characters | | 4, 688, 000 |
| Total bits | | 2.8 x 10 [7] |

D. *Use*

| | |
|---|---|
| Number of uses per minute [8] | 18. 3 |

[1] Estimate from Stack and Reader Division statistics.

[2] Estimated at approximately equal to 1955–61 average in the absence of any discernible trends plus 10 percent for possible growth.

[3] Balance of charges (1,100,000—200,000=900,000) estimated to be no more than ⅓ staff by Planning Research Corporation and the Office of the Information Systems Specialist.

[4] Estimates by Planning Research Corporation roughly based on average length of loan with a margin added to account for possible delinquency and time required for reshelving.

[5] Estimates by Planning Research Corporation.

[6] See Exhibit 16, note (4).

[7] Annual volume times reciprocal of minimum annual turnover times peak load multiple.

[8] Identical with location directory.

The process by which frequency of use was estimated for this file is closely parallel to that used in the case of the central catalog (Function D). The number of uses per operation, however, was adjusted to reflect the differing needs for authority information in various library operations. Resulting estimates, which range from 11.4 to 13.8 uses per minute, are much lower than estimates of total catalog use, since many types of catalog use are only occasionally associated with simultaneous use of the catalog authority file.

## *The Combined Function Groups*

The preceding discussion stressed the interrelationship between the files associated with various functions of the Library. For example, the close relationship between the location directory and the not-on-shelf file is obvious, since they are always consulted simultaneously. The circulation file is also closely associated with these files, since circulation of material is reported to the not-on-shelf file, and all circulation presupposes a prior retrieval via the location directory. By relating storage files to function groups rather than to individual functions, certain economies in equipment purchase may be achieved, particularly with respect to the elimination of redundant support and searching equipment.

A similar association exists between the monograph processing file, the serial file, and the exchange record. In addition to the fact that they all record the acquisition or processing of new items for the Library, they are use-related because it may not be immediately evident in which file a particular title might be, since the distinction between monographs and serials is not always clear, and since both types of acquisitions need to be included in the exchange records.

EXHIBIT 22.—CIRCULATION FILE SPECIFICATIONS

A. *Initial size:*
Zero. File will begin operation empty.

B. *Annual volume:*
Number of loans,[1] 220,000.

C. *Required capacity:*

$$\text{Required capacity} = \frac{\text{Annual volume}}{\text{Minimum annual turnover}} \times \text{Peak load factor}$$

Minimum annual turnover [2]=10

Peak load factor [3]=1.5

$$\text{Required capacity} = \frac{220{,}000}{10} \times 1.5 = 33{,}000 \text{ loans}$$

Number of characters required:

| | |
|---|---|
| 33,000 item descriptions at 70 characters each [4] | 2,310,000 |
| 33,000 borrower numbers at 6 characters each [5] | 198,000 |
| 6,600 borrower descriptions [6] at 80 characters each [5] | 528,000 |
| | 3,036,000 |

Number of bits required: $1.8 \times 10^7$

D. *Use:*

| | |
|---|---|
| Annual volume of loans | 220,000 |
| Number of uses per year at 2.5 uses per loan [7] | 440,000 |
| Number of uses per hour with 3,000 hours per year | 146.7 |
| Number of uses per minute | 2.4 |

[1] [2] [3] See Exhibit 21 notes (2), (4), and (5), respectively.

[4] Estimate based on sum of following items: main entry (18 characters); title (40); date of imprint (4); item number (8).

[5] Estimates by Planning Research Corporation.

[6] Assumes average of 5 loans per borrower. This is a very conservative estimate, since many Government agencies borrow hundreds of items at a time.

[7] Assumes one use for entering all loans and an average of 0.5 recall uses for each item loaned. Canceling loans (one use per item) will be accomplished automatically with reading devices.

The third combination of storage files involves the central catalog and authority file complex. The latter file would usually be consulted prior to using the catalog, although once an entry had been determined subsequent uses of the catalog might be made directly. Undoubtedly a small portion of the total catalog use would involve simultaneous searching of both files.

The addition of the authority file to the central catalog increases the required capacity by less than 5 percent, which, in view of the proposed trillion-bit file, further enhances the economies of combination. In Exhibit 24, the combination of the various files into groups has been shown in tabular form. In the balance of this report other hardware requirements, as well as costs, will be estimated for these combined function groups rather than for the individual functions.

The hardware required for these groups consists of general-purpose computers, storage files, auxiliary tape transport units, and, in some cases, on-line and off-line input and output equipment. In this report "general-purpose computers and support equipment" designates a complete functioning computer installation, including at least one tape transport unit but excluding the large auxiliary storage files and certain specific off-line equipment that will be enumerated.

For the internal processing group, a medium-sized general-purpose computer was included, in addition to the storage file and two auxiliary tape transport units. An unusually large or sophisticated computer should not be required for this group, since neither the frequency of access (22.5 per minute) nor the complexity of the search routines requires unusual electronic sophistication. The auxiliary tape units are included to allow for the numerous special programs and readouts which would be generated from this file for billing, acknowledging, accounting, scheduling, budget predicting, etc. The other major hardware item is the storage file, which is adequately defined by the capacity specifications summarized in Exhibit 24.

The equipment complement for the circulation and retrieval group is similar to that for the internal processing group. The combined storage files could be adequately served by a single medium-sized computer for the same reason cited above, although this file has a larger capacity and would be used more frequently. This group would also require auxiliary tape transports and a storage file unit.

In addition to the major hardware items discussed above, a large number of on-line input and output devices would be required. As outlined in Section III, status changes for circulated items would be reported semiautomatically by devices capable of reading magnetic print and reporting changes to the not-on-shelf file. Similarly, requests for retrieval would be relayed to the appropriate stack locations in the form of automatically typed labels, which could be machine-interpreted. Approximately 50 automatic typewriters and 70 magnetic-print readers would be required in order to provide complete coverage for the stack decks and the various points in the Library where circulated materials are charged or returned.

The central catalog group differs from the other combined function groups in that its specifications cannot be stated in terms of presently existing equipment, although a general outline of what is expected of the automated search logic and catalog is given in Section III. A more detailed exposition of the specific hardware requirements is not possible without more study and statistical analyses of the number and type of possible and/or probable search strategies expected in the automated system. In view of this, the storage file and related hardware are treated as a unit; cost estimates are based on overall costs suggested by industry representatives familiar with analogous existing systems and with development efforts.

The general specifications for the two types of consoles are for the most part identical, since they both require a maximum number of preset message formats and addresses, have similar keyboards, and require display screen capacity of a minimum of 1,000 characters, with equivalent internal buffering. Detailed specifications of the minimum console keyboard character set are presented in Exhibit 25.

EXHIBIT 23.—CATALOG AUTHORITY FILE: SPECIFICATIONS

I. *Initial size*

| Type of entry | Number of entries | Characters per entry | Total characters |
|---|---|---|---|
| Author (personal, corporate, form heading, etc.) | [1] 2,490,000 | [4] 68 | 169,320,000 |
| Subject headings | [2] 114,000 | [4] 23 | 2,622,000 |
| Subject cross-references | [3] 171,000 | 23 | 3,933,000 |
| Cataloging rules | 1 | [5] 2,890,000 | 2,890,000 |
| Total characters | | | 178,765,000 |
| Total bits | | | 1.1x10[9] |

II. *Maximum capacity*

Estimated at twice 1972 size ..... 2.1x10[9] bits

III. *Staff use*

| Operation | Number of uses per operation [6] | Number of uses per year |
|---|---|---|
| Establish authorities (61,200) [7] | 3 | 183,600 |
| Change authorities (12,900) [8] | 2 | 25,800 |
| Establish subject headings (4,100) [9] | 2 | 8,200 |
| Change subject headings (200) [10] | 3 | 600 |
| Cataloging and recataloging (151,000) [11] | 2 | 302,000 |
| Editing new NUC titles (118,000) [11] | 1 | 118,000 |
| Reference correspondence (143,000) [11] | 0.2 | 28,600 |
| Other staff use (3,699,000) [11] | .1 | 369,900 |
| Total staff use per year | | 1,035,800 |
| Staff use per hour for 3,000 hour year | | 342.4 |

IV. *Summary: All uses*

| | *Per hour* |
|---|---|
| Staff use | 342.4 |
| Reader and in-person reference use: [12] | |
| Present level | 237.3 |
| Increase of 50 percent | 355.9 |
| Card Division author/title or subject searches [12] | 101.6 |
| Regional network use: [12] | |
| Low forecast | 10 |
| High forecast | 40 |

See footnotes at end of table.

*Total use per minute*

| | Reader use projection | |
|---|---|---|
| | Constant | Up 50 percent |
| Low regional use forecast | 11.5 | 13.5 |
| High regional use forecast | 12.0 | 14.0 |

[1] Estimate derived in the following manner: average number of catalog cards per personal or corporate author in the Main Catalog, according to study made by the Information Systems Specialist in 1961, was 5.9. Applying this figure to the total volume of cards in the Main Catalog yields an estimate of 1,852,000 authors. To this estimate was added the projected number of authorities to be established 1962–72 which equals 633,600 according to a linear projection, for which the corresponding coefficient of correlation is .90.

[2] Based on the sum of the following:

| | | |
|---|---|---|
| Subject headings June 4, 1953. | 48, 833 | Subject Cataloging Division |
| Headings added FY '54–FY '61. | 21, 520 | Library *Annual Reports* |
| Headings to be added FY '62–FY '72. | 43, 890 | Derived from linear projection based on 1953–61 data (r=.89) |
| Total | 114, 243 | |

[3] Based on ratio of 1.5 cross-references per subject heading according to Subject Cataloging Division estimates.

[4] Number of characters per author (18) and per subject heading (23) from Exhibit 14. An additional 50 characters per author are estimated for information to establish the authority.

[5] Based on estimated contents of the Library's *Rules for Descriptive Cataloging* (1959) and the American Library Association's *A. L. A. Cataloging Rules for Author and Title Entries* (1949).

[6] Estimated by Planning Research Corporation.

[7] Based on linear regression analysis of the years 1953–1961, with a corresponding correlation coefficient (r=.90).

[8] Based on linear regression analysis of the years 1949–1961, with a corresponding correlation coefficient (r=.89).

[9] Based on linear regression analysis of the years 1949–1961, with a corresponding correlation coefficient (r=.89).

[10] Assumed constant at rate equal to 1949–1961 average.

[11] See Exhibit 14, total staff use minus cataloging, recataloging, editing new NUC titles, and reference correspondence.

[12] Estimated to be .1 of total catalog use as calculated in Exhibit 14.

EXHIBIT 24.—COMBINED FUNCTION GROUPS

| | *Capacity (bits)* | *Frequency of access (uses per minute)* | |
|---|---|---|---|
| *Internal processing group:* | | | |
| A. Monograph acquisition and processing | $1.6 \times 10^8$ | 5.3 | |
| B. Serial acquisition, processing, and holdings | $5.3 \times 10^8$ | 14.4 | |
| P. Exchange record | $2.0 \times 10^8$ | 2.8 | |
| Total for combined group | $8.9 \times 10^8$ | 22.5 | |
| *Central catalog group:* [1] | | | |
| D. Catalog and automated search logic | $4.5 \times 10^{10}$ | [2] 194.0 | [3] 124.2 |
| C. Catalog Authority File (authors, subjects, cataloging symbols, rules, etc.) | $0.2 \times 10^{10}$ | [2] 14.0 | [3] 11.5 |
| Total for combined group | $4.7 \times 10^{10}$ | [2] 208.0 | [3] 135.7 |
| *Circulation and retrieval group:* [4] | | | |
| G. Control of item handling and location: | | | |
| Location directory | $11.7 \times 10^8$ | 18.3 | |
| Not-on-shelf file | $0.28 \times 10^8$ | 18.3 | |
| I. Circulation file | $.18 \times 10^8$ | 2.4 | |
| Total for combined group | $12.16 \times 10^8$ | 39.0 | |

[1] This group also includes Function N (Monitoring catalog use and response) for which no separate specifications were developed.

[2] Presumes high forecast of regional use and a 50 percent increase in reader use.

[3] Presumes low forecast of regional use and no change in reader use.

[4] This group also includes Function M (Monitoring use of the collection) for which no separate specifications were developed.

The various types of preset message formats and predetermined operations for the consoles are described in Exhibit 26. Not all of these formats or operations would be relevant to all consoles; the output-only consoles, for instance, would require three basic instructions: display, hard copy, and retrieve. In addition to these instructions, the input-output consoles would have "enter" and "erase" instructions. A description of the console variations for various Library operations is presented in part C, Exhibit 26.

In the system used for the costing analysis there are 9 possible addresses or locations to which a console query may be directed and 8 storage files which may be queried: the central catalog, authority file, monograph processing file, serial file, location directory, not-on-shelf file, circulation file, and the exchange record. In addition, queries or instructions could be sent to the editing processor. Since several of these files are relevant only to specific operations in the Library, the maximum number of addresses that any specific console would require is 6.

EXHIBIT 25.—CONSOLE KEYBOARD CHARACTER SET

| *Category* | *Minimum number of characters* |
|---|---|
| English alphabet (upper and lowercase) | 52 |
| Special characters in other Roman alphabets (upper and lowercase)[1] | 16 |
| Numbers | 10 |
| Punctuation marks[1] | 19 |
| Diacritical marks[1] | 14 |
| Commonly used symbols[1] | 6 |
| Field symbols[2] | 11 |
| Total | 128 |

[1] Based on actual count in the GPO *Style Manual* (1959), which describes the following languages using the Roman alphabet:

| | |
|---|---|
| English | Latin |
| Danish | Norwegian |
| Dutch | Portuguese |
| French | Spanish |
| German | Swedish |
| Hungarian | Turkish |
| Italian | |

[2] Includes only the most frequently used search modes. Numerous other combinations could be provided with nonsense letter configurations. The included fields are:

| | |
|---|---|
| Author | Pagination |
| Title | Edition statement |
| Subject | Illustration statement |
| Date | Item number |
| Language | Classification number |
| Country of publication | |

The probability that a significant part of the total cost for the central catalog would include presently existing support equipment is not too great, since much of the input and many of the specialized outputs would be controlled by means of hardware designed to edit input and output automatically (see Function J). Similarly, any extensive program for modifying or reproducing the central catalog would also operate through this editing processor.

An additional function to be implemented by the basic hardware for the automated catalog is N, Monitoring Catalog Use and Response. This monitor should involve only a limited number of recording routines within the automated search logic and partial use of a tape transport for periodic readout and tabulation. This function is essentially internal to the automated catalog, and therefore hardware requirements and cost are not analyzed separately.

## *Consoles*

With the installation of a central catalog in digital form, the equipment required for access to this catalog and subsidiary files becomes of paramount importance in the Library operation. A large number of consoles would be required to provide the necessary access to these files for the Library staff and the readers. Although these consoles will vary somewhat, depending upon the particular files and operations with which they are associated, they are essentially of two types: the input-output console and the output console. The input-output console, for staff use only, permits adding to or changing the records in the various storage files. The output console is the more general type and would be used by both staff and readers for purposes of querying the central catalog and subsidiary files.

Determination of the number of consoles required throughout the Library poses several problems. In many instances it was only possible to estimate very approximately the number of console operations connected with a specific activity. Wherever possible, the volume of such console use was taken at the high end of the probable range of estimates. In some library operations the use of the catalogs and files, which in the automated system would require a console, is peripheral or occasional. In such cases, the number of consoles required was based on the number of man-years estimated to be spent, at present levels of operation, in catalog use. The resulting estimate

EXHIBIT 26.—PRESET MESSAGE FORMATS AND OPERATION INSTRUCTIONS FOR THE CONSOLES

A. *Message formats*

Input formats:

a. Central catalog: simplified format for most entries.
b. Central catalog: extended format for special or complex entries.
c. Catalog authority file.
d. Monograph processing file.
e. Serial file.
f. Location directory.
g. Circulation file.
h. Exchange file.

Posting formats:

i. New holdings reported to the National Union Catalog.
j. Checking in new issues in serial file.
k. Status changes for monograph processing file, serial file, and exchange file.[1]

Search formats:

l. Basic format for all bibliographic files.[2]
m.–n. Specialized formats (two) for more sophisticated searching of central catalog.[3]

Special formats:

o. Search circulation file by borrower.
p. Instruct editing processor.

B. *Operation instructions*

1. Display.
2. Printout (tear-off hard copy).
3. Retrieve.
4. Enter.
5. Erase.

C. *Message formats and operations per console:*

1. *Type and use of console:*

| | Message formats used | | Operation instructions | |
|---|---|---|---|---|
| | Total number of formats | (See A above) | Total | (See B above) |
| All output-only (Reader) consoles | 3 | l–n | 3 | 1–3 |
| Input-output (Staff) consoles: | | | | |
| Cataloging | 8 | a–c, h, k, l–n | 5 | 1–5 |
| Monograph processing | 5 | d, k, l–n | 5 | 1–5 |
| Serial processing | 6 | e, j, k, l–n | 5 | 1–5 |
| Exchange and gift | 5 | h, k, l–n | 5 | 1–5 |
| Circulation control | 5 | g, l–n, o | 5 | 1–5 |
| Editing (input and publications) | 4 | l–n, p | 5 | 1–5 |
| Item control | 4 | g, l–n | 5 | 1–5 |

2. *Maximum per console:*

Output-only consoles:

3 formats, 3 operational instructions.

Input-output consoles:

8 formats, 5 operational instructions.

[1] Refers to such limited posting operations as noting item ordered, received, or canceled; item offered, withdrawn, received; serial title discontinued.

[2] Includes author, title, subject, item number, and classification number search modes.

[3] Provides for searching by less frequently used modes, e.g., NUC library symbol, publisher, date, pagination, etc., and by special combinations of search modes, e.g., author-date-language, date-pagination-subject, subject-illustration-date, etc.

of man-years was adjusted to allow for inereased effieiency of console use compared to manual eonsultation of files.

Exhibit 27 indicates the estimated number of eonsoles required by various library operations, categorized by input-output consoles and by output consoles. Beeause of the problems in estimating probable use of the system by readers and by outlying libraries via the regional communications network, consoles for these uses were estimated parametrieally and therefore expressed as a range. The maximum number of eonsoles antieipated is 409 for all uses; the minimum is 294. In both instanees the number of input-output eonsoles is 55. Excluding fee serviees and regional uses, the number of eonsoles used in the Library is either 148 or 163 depending on reader use.

EXHIBIT 27.—NUMBER OF CONSOLES REQUIRED

| Category | Division or function | Number of consoles | | |
|---|---|---|---|---|
| | | Input-output | Output only | Total |
| A. Staff use | 1. Serial acquisition and processing | 15 | 1 | 16 |
| | 2. Cataloging [1] | 5 | 20 | 25 |
| | 3. Monograph acquisition and processing | 5 | 1 | 6 |
| | 4. Location directory [2] | 2 | 0 | 2 |
| | 5. Exchange file | 3 | 0 | 3 |
| | 6. Editing Processor | 1 | 0 | 1 |
| | 7. Circulation file [3] | 8 | 0 | 8 |
| | 8. NUC posting | 12 | 1 | 13 |
| | 9. Reference service | 0 | 22 | 22 |
| | 10. Searching for recommending function (Reference Department and Law Library). | 0 | 2 | 2 |
| | 11. Selection Officers | 2 | 0 | 2 |
| | 12. Cyrillic | 2 | 0 | 2 |
| | 13. Acrospace Information Division [4] | 0 | 3 | 3 |
| | 14. Defense Research Division [4] | 0 | 2 | 2 |
| | 15. Legislative Reference Service | 0 | 11 | 11 |
| B. Reader use | Reader (present levels of operation) | 0 | 30 | 30 |
| C. Fee service use | 1. Photoduplication Service | 0 | 2 | 2 |
| | 2. Card Division | 0 | 44 | 44 |

SUMMARY

| Broad categories of use | Number of consoles | | |
|---|---|---|---|
| | Input-output | output only | Total |
| The basic Library (includes Staff, Reader, and LRS) | 55 | 93 | 148 |
| Fee services (Card Division and Photoduplication Service) | 0 | 46 | 46 |
| Regional communication terminals (low forecast) | 0 | 100 | 100 |
| Minimum total | 55 | 239 | 294 |
| Adjustment for 50 percent increase in reader use | 0 | +15 | +15 |
| Adjustment for high forecast of regional communication terminals | 0 | +100 | +100 |
| Maximum total | 55 | 354 | 409 |

[1] Includes editing of new NUC titles reported for more complete integration into central catalog. (This does not include posting additional locations symbols, see 8 under Staff Use.)

[2] Location directory only. No allowance is made for NOS file where all posting will be automatic.

[3] Includes searching done by the Loan Division for some prospective borrowers.

[4] Funded by Department of Defense, administered by the Library of Congress.

## *Hardware Requirements for Editing and Publication Functions*

The term "editing processor" is used in this report to designate the hardware equipment associated with Function J, Editing for Input and Output. As outlined in Section III, this function is concerned both with editing input data for final inclusion in the catalog and also for the selection and editing of entries for publications based on data in the central catalog. If a read-only central memory were selected for the automated Library, then the major hardware item required for this function would be a large general-purpose computer capable of readily selecting, editing, and processing both input and output from the catalog. In addition to the general-purpose computer and auxiliary storage file, it would be necessary to provide for frequent additions and changes, since the storage file associated with the editing processor would consist almost exclusively of detailed instructions for the more common operations, dictionaries of prohibited and key operational terms, etc., and since this file would probably not be updated very often. In this respect the storage file associated with the editing processor resembles the central catalog which would probably be amended and merged at regular intervals, although not as frequently as the processing and circulation files.

The type of file best suited to the infrequently amended dictionary portion of the editing processor and also to the automated catalog is often called an "integral search" file. This type of file (frequently in read-only form) can be amended and modified, but it may be less well suited to the continual changes associated with recording frequent transactions. Read-only files integrally addressed, however, can be searched at very high speeds, and are, therefore, particularly attractive for operations involving the central catalog.

When the editing processor has selected, arranged, and edited catalog information for publication in bibliographies, accessions lists, etc., the resulting output is a digital form record (e.g. magnetic tape) with coded instructions on format and typeface. This digital output is converted to printed page masters for high quality reproduction of cards, catalogs, bibliographies, and indexes. The primary hardware associated with this operation (Function K) is a large photocomposition device. This photocomposer must be capable of producing all Library lists, bibliographies, and catalogs, and possibly some special-purpose bibliographies produced on a fee basis for patrons.

This publication output, as measured in words or pages, varies sharply from year to year depending on the number and size of special bibliographies and reports. A brief survey of Library publications made by the Office of the Information Systems Specialist indicates that the 12 largest recurrent publications totaled approximately 19 million words annually. Without further detailed study of the special publications, no systematic projection of future, or even present, annual volume is possible. Similarly, no estimate can be made of the demand for special-purpose bibliographies on a fee service basis without a study of possible fee schedules and some analysis of the market for this service.

For the purpose of cost analysis a 50-million word-per-year estimate was used as an approximation of the photocomposer workload. To handle such a load, the photocomposer must have an output rate of about 100 characters per second. An examination of the most important recurring publications leads to the reasonable but very rough estimate that, to produce similar output, the photocomposer would need 20 typefaces in no less than 5 different point sizes.

It is assumed that this publishing activity would be relatively flexible and could be carried on in off hours to avoid peak loading problems during the hours of public service.

## *Hardware for Graphic Storage and Retrieval*

There are many collections in the Library in which there is a possible need for special graphic storage and retrieval equipment, e.g. maps, prints, photographs, and manuscripts. Although these may all be considered as possible future applications, only one such application was included in the cost analysis: the current reference file maintained by the Legislative Reference Service. This file contains approximately 500,000 documents and clippings, each of which remains in the file for 5 years; thus about 100,000 documents are purged and added each year, keeping the file roughly at a constant size.

Graphic storage equipment would permit extensive cross-referencing and would insure file integrity while reducing search time. These objectives can be accomplished with digitally coded high-speed microfilm storage. Other requirements include production of hard copy in a few seconds and also the rapid merging, reorganization, and purging of the file to keep it timely and relevant.

The Office of the Information Systems Specialist estimated that this file is used approximately 870 times per week, or just under 22 uses per hour, based on a 40-hour week. Since this file must be readily available at all times to virtually the entire staff of the Legislative Reference Service, a peaking factor of approximately 3 was used to insure minimum queuing.

The hardware equipment associated with this function includes an input processing device that combines the feature of a microfilm camera and a Flexowriter (or similar device) for imprinting the film with the appropriate digital code, and at least two complete retrieval units for film display and rapid hard-copy reproduction. The equipment should be capable of producing microfilm copy rapidly, in order to facilitate merging and multiple filing.

## *Other Hardware Requirements*

Two other functions for which hardware is required are O, Business Data Processing and Q, Scheduling Library Operations. The former will not be considered in detail because automation of the Library's accounting routines is now underway and the present tabulating equipment will be replaced by a small computer in 1964. However, the cost of this computer and related equipment is included in the cost analysis. It is anticipated that the scheduling of Library operaions should require no more than occasional use of the same small general-purpose computer to produce analyses and data projections of operations throughout the system. Consequently, requirements and costs for this function are not treated in detail.

## *Functions Excluded or Analyzed Separately*

Several functions listed in Section III have not been discussed with respect to hardware requirements. For example, Function M, Monitoring Use of the Collection, was excluded because of its relative simplicity and close association with other functions. Since each use is recorded in the not-on-shelf file, the associated computer can easily be programmed to store the information for a use record which may be read out at frequent intervals and stored on magnetic tape for further analysis.

Function H, the Physical Storage of Items, requires no specific hardware other than that associated with the not-on-shelf file and the location directory. Automatic typewriters and mechanical reading devices will be used throughout the stack areas to permit control of items. Although it is possible to envision library stacks as huge automated warehouses with machines which automatically select desired books from the shelves, these possibilities were not included in this analysis. Thus the hardware implications of physical storage of the collection are essentially potential rather than actual insofar as more efficient modes of storage and retrievel may be developed once the automated system is installed.

The Production of Photocopy, Function L, was not included in the cost analysis. The Library presently maintains a large photoduplication laboratory with extensive photocopy and microfilm equipment, and it is not anticipated that the automated system would immediately modify or replace this equipment.

The hardware associated with regional transmission, Function S, consists of output consoles in the Library and a communications network. The particular mode of transmission to be used has not been determined and was not considered of particular relevance to this analysis, since transmission costs will undoubtedly be borne, at least in part, by other libraries, and the communications network will be developed as a joint effort; hence, it is not truly a part of the internal system cost of the Library.

Function T, Converting Existing Files to Machine-Readable Form, presents alternatives with respect to cost, since the function may be considered as a continuing operation of the automated system or as a unit process to be completed before the system becomes fully operational. These alternatives and some relevant cost facts are treated in Section VIII.

The digital-form input processing (Function U) of the *text* of published material, as distinguished from bibliographic information, is contingent on the adoption of digital processing in the publishing industry, the successful development of economical and adaptable automatic conversion hardware, or the widespread use of linotype paper tapes for printing. Since the rate of technological progress in these areas cannot be predicted with any accuracy, no specific requirements are projected for Function U. Even if the necessary equip-

ment becomes available before the installation of the automated system, this should represent no significant increase in the total system cost and may, in fact, produce a net saving.

Many of the hardware requirements of digital-form input processing have been included in this analysis, e.g., those relating to storage capacity, automated search methods, and suitable output equipment. The most significant hardware cost not included would be the specific equipment required for automatic conversion of printed material to machine-readable form. The costs of this equipment may well be compensated for by the savings resulting in Library operations such as cataloging and other input operations.

The system displayed in Exhibit 1 shows the various hardware components communicating through an intercommunication switching network (Function V). The approximate number of terminals and individual circuits can be determined from the discussion on the number of location of consoles in the system. Further characterization of this network must await a detailed system engineering study.

## *Résumé of Hardware Specifications*

Exhibit 28 is an outline summarizing hardware specifications by function. Storage files associated with a particular function are related to the particular exhibit in which file requirements are displayed. In addition to the hardware indicated in Exhibit 28, two tape transport units and a high-speed printer are used in conjunction with several functions.

Hardware specifications in this section have largely been analyzed with respect to the individual functions. As mentioned previously, this does not always provide the most suitable unit for analysis of requirements or costs. As a consequence, various related functions have been combined into basic hardware groups. Exhibit 24 is a reference table assigning functions to the three main hardware groups. In the next two sections of this report, costs will be developed for these three groups as well as for specific hardware required to automate single functions.

EXHIBIT 28.—SUMMARY OF HARDWARE SPECIFICATIONS AND REQUIREMENTS BY FUNCTION

*Function A.* Recording monograph acquisitions and processing.[1]
- Associated storage file: Monograph processing file (Exhibit 17).
- File capacity: $1.6 \times 10^8$ bits.
- Frequency of use: 5.3 per minute.

*Function B.* Recording serial acquisition, processing, and holdings.[1]
- Associated storage file: Serial file (Exhibit 18).
- File capacity: $5.3 \times 10^8$ bits.
- Frequency of use: 14.4 per minute.
- Hardware requirements: Subsumed under internal processing group (see Function A).

*Function C.* Maintaining catalog authority files.[2]
- Associated storage file: Dictionary file of authors, subject headings, etc. (Exhibit 23).
- File capacity: $6.5 \times 10^8$ bits.
- Frequency of use: 11.5–14 per minute.

*Function D.* Catalog use.[2]
- Associated storage file: Central catalog (Exhibits 15 and 16).
- File capacity: $4.5 \times 10^{10}$ bits.
- Frequency of use: 124.2–194 per minute.

*Functions E and F:*
- E—File access for staff (via input/output consoles).
- F—File access for readers (via output consoles).
  - Screen capacity and internal buffering (minimum): 1,000 characters.
  - Keyboard character set (minimum): 128 characters.
  - Maximum number of pre-set message formats: 8.
  - Maximum number of addresses: 7.
  - Maximum number of operative instructions: 5 (input-output), 3 (output-only).
  - Number of consoles required:
    - Minimum: 294 (including 55 input-output).
    - Maximum: 409 (including 55 input-output).

*Function G.* Control of library holdings.[3]
- Associated storage files: Location directory (Exhibit 20). Not-on-shelf file (Exhibit 21).
- File capacity (bits): Location directory $1.2 \times 10^9$. Not-on-shelf file $2.8 \times 10^7$.
- Frequency of use: 18.3 per minute.

*Function H.* Physical storage of items.

No requirements or specifications are to be provided for this function except as a part of the relocation operations associated with Function G above.

*Function I.* Circulation control.[3]
- Associated storage file: Circulation file (Exhibit 22).
- File capacity: $1.8 \times 10^9$ bits.
- Frequency of use: 2.4 per minute.

*Function J.* Editing for input and output.
- Hardware requirements:
  - 1 fast-access storage file of $2.5 \times 10^8$ bits.
  - 1 large general-purpose computer.
  - 10 tape transport units.

[1] This function is included in the internal processing group for which specifications are given at the end of the exhibit.

[2] This function is included in the central catalog group for which specifications are given at the end of this exhibit.

[3] This function is included in the circulation and retrieval group for which combined specifications are given at the end of this exhibit.

*Function K.* Publication of bibliographic materials.

Hardware requirements:

1 large automatic digitally controlled photocomposition device (photocomposer).
100 characters per second.
20 type faces in five point sizes.
5 basic publication formats.

*Function L.* Production of photocopy.

No hardware requirements are specified for this function beyond the present hardware complement in the Library.

*Function M.* Monitoring use of the collection.

No separate specifications are given, since this function will be a programmed routine of the general-purpose computer associated with the circulation and retrieval group.[4]

*Function N.* Monitoring catalog use and response.

No separate specifications are given, since this function will comprise a set of routines and off-line recording associated with the central catalog group.[4]

*Function O.* Business data processing.

No capacity specifications are derived in this report, since the Library is presently planning installation of data processing equipment.

Hardware requirements:

1 small general-purpose computer.
1 high-speed printer.
1 tape transport.

*Function P.* Maintaining exchange and gift records.[5]

Associated storage file: Exchange record (Exhibit 19).
File capacity: $2 \times 10^5$ bits.
Frequency of use: 2.8 per minute.

*Function Q.* Scheduling Library operations.

No separate specifications are given. It is expected that this function will make use of equipment associated with Function O for its electronic data processing needs.

*Function R.* Providing special graphic retrieval.

Associated storage file: Legislative Reference Service clipping file.
File capacity: 500,000 items (page units).
Frequency of use: Estimated peak, 60 per hour (average, 22 uses per hour).
Hardware requirements:

1 recording unit.
1 1:1 camera.
3 retrieval units.
2 paper tape read-and-punch typewriters.

*Function S.* Communication with other institutions.

Hardware requirements:

100 output-only consoles (low forecast).
200 output-only consoles (high forecast).
Necessary communication lines and terminal sets.

No more refined specifications are necessary for purposes of this report, since the costs of hardware and communications lines associated with this function will be borne by the other libraries that use this service and will therefore not be a part of the system cost to the Library of Congress.

*Function T.* Conversion of existing files to machine-readable form.

Cost and requirements of this function are treated separately in Section VIII.

*Function U.* Digital-form processing.

Requirements and costs for this function are not treated in this report, since it is premature to estimate the availability of published material in digital form or the necessary equipment for internal conversion to digital form.

*Function V.* Controlling intrasystem communication.

No specific requirements are given. This function will be costed as a unit.

### Summary

*Basic Hardware Groups*[6]

Processing group:

Storage file capacity: $8.9 \times 10^8$ bits.
Frequency of use: 22.5 per minute.
1 medium-size general-purpose computer.
3 tape transport units.

Central catalog group:

Storage file capacity: $4.7 \times 10^{10}$ bits.
Frequency of use: 135.7–208 uses per minute.
1 trillion-bit file with automated search logic.

Circulation and retrieval:

Storage file capacity: $1.2 \times 10^9$.
Frequency of use: 39.0 per minute.
1 medium-size general-purpose computer.
3 tape transports.
50 off-line electric typewriters.
70 limited-capacity magnetic print readers adapted for item control.

[4] Combined specifications for this group are given at the end of this exhibit.

[5] This function is included in the circulation and retrieval group for which combined specifications are given at the end of this exhibit.

[6] See Exhibit 24 for tabulation of included functions, capacity, and frequency of use.

SECTION VI

# *Hardware and Software Costs*

## *Summary*

The total hardware costs of the automated system are estimated at $16.7 million; software costs total about $9.6 million. The major hardware items in the total initial system costs of $26.3 million are a $5 million central file unit and 163 consoles expected to cost about $5.4 million. This report assumes the purchase of equipment rather than rental.

## *Hardware Costs*

In this section procurement and software costs are developed for the hardware in each of the functional groupings discussed in the previous section. The basic hardware costs are presented in Exhibit 29, and their relationship to the functions discussed in Sections III and V is shown. Whenever possible, the costs of hardware items are taken directly from published pricelists. Obviously, in some instances there is no presently available comparable equipment—for example, the central memory, the photocomposer, and the peripheral equipment used in recording circulation status. In these cases, however, somewhat related equipment is either under development or considered feasible, based on known advanced technologies. Both the proposed trillion-bit memory and the required photocomposer are essentially extensions of existing technology. Research that may lead toward the development of these items is proceeding independently of the potential automation of the Library of Congress. Estimates for all hardware under development were derived from discussions with industry representatives.

In Exhibit 29 hardware items are identified with arbitrary numbers, since it is not the purpose of this report, nor the intention of the survey team to recommend any manufacturer or equipment. Costs presented are suggested as guidelines rather than precise bids or offers by any company.

Where prices are available, the costs associated with basic hardware groups are detailed by component. Most groups include three major types of components: general-purpose computers, digital storage files, and additional tape transport units. The hardware complexes for which the highest expenditures are anticipated are the central catalog group, the consoles, and the editing processor. The cost of the central catalog group is estimated at at least $5 million, since its main component, the trillion-bit memory, is still under development. The cost of the consoles for the Library system is $5,440,000, the estimated cost for 163 consoles including software. Only consoles that are a direct burden of the Library budget are included on the assumption that consoles used in conjunction with fee services will be purchased through the receipts of these services, and consoles at regional terminals throughout the country will be paid for by the user libraries. The last major hardware group, the editing processor, is estimated to cost $3,362,500. Other hardware groups in the system range in cost from $120,000 (business data processing and scheduling) to almost a million dollars for the circulation and retrieval complex.

Exhibit 30 indicates the sources for the cost estimates and lists manufacturer designations for equipment items that meet the requirements of the system hypothesized for costing purposes. Since comparative costs and specifications are most readily available for general-purpose computers, at least three different computers are presented for comparison. Peripheral equipment prices are highly competitive, but since these costs are relatively small with respect to the total systems costs, the most readily available price lists were used. The special reference graphic storage equipment estimates are based on quoted prices by the FMA, Inc. (Filesearch system), without implication by the survey team that competitive equipment does or does not exist for this task.

EXHIBIT 29.—BASIC HARDWARE COSTS OF THE AUTOMATED SYSTEM

| *Basic Hardware Group* | *Item number* [1] | *Purchase price* [2] |
|---|---|---|
| 100 *Internal Processing Group* | | |
| Functions included: | | |
| A. Recording monograph acquisition and processing. | | |
| B. Recording serial acquisition, processing, and holdings. | | |
| P. Maintaining exchange and gift records. | | |
| Basic hardware items: | | |
| 1 medium-size general-purpose computer and support equipment [3] | 101 | $282, 750 |
| 1 digital storage file (capacity $\geq 9 \times 10^8$ bits) | 102 | 388, 500 |
| 2 additional tape transport units | 103 | 72, 500 |
| Total cost | .... | 743, 750 |
| 200 *Circulation and Retrieval Group* | | |
| Functions included: | | |
| G. Control of library holdings. | | |
| H. Physical storage of items. | | |
| I. Circulation control. | | |
| M. Monitoring use of the collection. | | |
| Basic hardware items: | | |
| 1 medium-size general-purpose computer and support equipment [3] | 201 | 282, 750 |
| 1 digital storage file (capacity $\geq 1.3 \times 10^9$ bits) | 202 | 504, 500 |
| 50 on-line electric typewriters at $800 | 203 | 40, 000 |
| 70 limited-capacity magnetic-print readers at $1,000 | 204 | 70, 000 |
| 2 additional tape transport units | 205 | 72, 500 |
| Total cost | .... | 969, 250 |
| 300 *Central Catalog Group* | | |
| Functions included: | | |
| C. Maintaining catalog authority files. | | |
| D. Catalog use. | | |
| N. Monitoring catalog use and response. | | |
| Basic hardware items: | | |
| 1 trillion ($10^{12}$) bit file and support equipment | 301 | 5, 000, 000 |
| Total cost | .... | 5, 000, 000 |
| 400 *Consoles* | | |
| Functions included: | | |
| E. File access for staff | | |
| F. File access for readers | | |
| Basic hardware items: | | |
| 55 input-output (staff consoles) | 402 | 2, 200, 000 |
| 108 output-only (reader) consoles [4] | 402 | 3, 240, 000 |
| Total cost | .... | 5, 440, 000 |
| 500 *Editing Processor* | | |
| Function included: | | |
| J. Editing for input and output | | |
| Basic hardware items: | | |
| 1 large general-purpose computer with support equipment [3] | 501 | 2, 500, 000 |
| 1 very fast access digital storage file (capacity $\geq 2.5 \times 10^8$ bits) | 502 | 500, 000 |
| 10 additional tape transport units | 503 | 362, 500 |
| Total cost | .... | 3, 362, 500 |

See footnotes at end of table.

| *Basic Hardware Group*—Continued | *Item number* [1] | *Purchase price* [2] |
|---|---|---|
| 600 *Photocomposer* | | |
| Function included: | | |
| K. Publication of bibliographic materials | | |
| Basic hardware items: | | |
| 1 automatic photocomposition device capable of producing 50 million words per year with 20 different typefaces | 601 | 500, 000 |
| Total cost | | 500, 000 |
| 700 *Business Data Processing and Scheduling* | | |
| Functions included: | | |
| O. Business data processing | | |
| Q. Scheduling Library operations | | |
| Basic hardware items: | | |
| 1 small general-purpose computer with support equipment [3] | 701 | 85, 000 |
| 1 high-speed printer | 702 | 35, 000 |
| Total cost | | 120, 000 |
| 800 *Special Reference Graphic Storage* | | |
| Functions included: | | |
| R. Providing special graphic retrieval | | |
| Basic hardware items: | | |
| 1 recording unit | 801 | 15, 000 |
| 1 1:1 camera | 802 | 7, 500 |
| 2 Flexowriters at $4,000 | 803 | 8, 000 |
| 3 retrieval units at $103,000 | 804 | 309, 000 |
| Total cost | | 339, 000 |
| 900 *Miscellaneous General Hardware* | | |
| Basic hardware items: | | |
| 1 high-speed printer | 901 | 35, 000 |
| 2 tape transport units (spares) | 902 | 72, 500 |
| 1 intrasystem communications network | 903 | 100, 000 |
| Total cost | | 207, 500 |

[1] Arbitrary numbers provided for reference within this report.

[2] Sources of rental prices by hardware item number are listed in Exhibit 30.

[3] Support equipment includes, among other items, a file control unit and one tape transport unit.

[4] The number of output-only (customer) consoles used excludes consoles associated with fee services or regional communication terminals. The cost of these consoles will probably be borne by customers for these fee services, and, similarly, regional network consoles will be paid for by user libraries. However, average unit costs are based on a projected production run that includes regional and fee service consoles.

## *Software Costs*

The costs of developing the necessary compilers and programs vary widely depending on their complexity and number, therefore software costs, unlike hardware costs, cannot be based on unit prices. In the computer industry it is generally estimated that the total software cost roughly equals the total hardware cost, a rule of thumb used here only as a general guide. Where applicable, software costs are estimated to be equal to the costs of programmable or logic-containing hardware. For this purpose, hardware is defined as including storage files; however, consoles, additional tape transport units, on-line reading devices, and automatic typewriters are excluded, since these items require no programming, once they are developed.

Attributing software costs to specific hardware groups can be misleading, even if the total software cost estimate is valid. This is particularly true with respect to the software costs of the central catalog group and the editing processor. Since the editing processor is closely connected with the central catalog, the software costs of these two groups must essentially be considered as a unit. In order to estimate the total amount of such costs, however, they have been separated and attributed to the individual groups in Exhibit 31.

The great difference between hardware ($16.7 million) and software costs ($9.6 million) is primarily atttributable to the high total cost of consoles. Although the use of consoles will require programming of certain subroutines in the associated general-

EXHIBIT 30.—SOURCES OF PRICE ESTIMATES FOR BASIC HARDWARE ITEMS

| Hardware item No.[1] | Manufacturer and model [2] | Monthly rental range[3] (dollars) | Comments |
|---|---|---|---|
| 101 | IBM 1410 | 6,000–32,000 | Medium-size general-purpose computer |
| | RCA 501 | 11,000–26,000 | |
| | Univac III | 15,000–30,000 | |
| 102 | IBM 1301 with file control | [4] 7,135 | Digital storage file, $3.3 \times 10^7$ bits per magnetic disk (3 disks required) |
| 103 | IBM 729/4 | [4] 900 | Tape transport unit |
| 201 | See item 101. | | |
| 202 | IBM 1301 with file control | [4] 9,235 | See item 102 (5 disks required) |
| 203 | Estimate [5] | | Automatic typewriters |
| 204 | Estimate [5] | | Magnetic print readers |
| 205 | See item 103. | | |
| 301 | Estimate [5] | | Trillion-bit file and search logic |
| 401 | Estimate [5] | | Input-output consoles |
| 402 | Estimate [5] | | Output-only consoles |
| 501 | IBM 7090 | 50,000–69,000 | Large general-purpose computer |
| | Univac 1107 | 32,000–60,000 | |
| | RCA 601 | 24,000–68,000 | |
| | Philco 2000 | 47,000–79,000 | |
| 502 | Photostore | [6] 11,000 | Very fast access digital storage file |
| 503 | See item 103. | | |
| 601 | Estimate [5] | | Photocomposer |
| 701 | IBM 1401 | 2,500–12,000 | Small general-purpose computer tentatively selected by the Library for business data processing |
| 702 | IBM 1403 | [4] 800 | High-speed printer |
| 801–804 | Based on prices provided by FMA, Inc., for components of their Filesearch system | | Special reference graphic storage |
| 901 | See item 702. | | |
| 902 | See item 103. | | |
| 903 | Estimate [6] | | Intrasystem communications network |

[1] Same item numbers used in Exhibit 14.

[2] Comparable equipment systems, all capable of meeting requirements of the relevant functions.

[3] From *Computer Characteristics Quarterly*, December 1961, Charles W. Adams Associates. Ranges reflect different peripheral equipment configurations available. Purchase prices estimated in Exhibit 14 were taken from actual pricelists where available, specifying needed peripheral equipment. All rentals are for 176 hours operation per month.

[4] Rental costs taken directly from published pricelists.

[5] No precise cost information was available for several items, either because they are still being developed (items 301, 401–2 and 601) or because they may require extensive modifications (items 203 and 204).

[6] The intrasystem communication network cost was roughly estimated from similar complex automated systems.

purpose computers, this cost is included in their estimated purchase cost. The balance of this discrepancy is accounted for by the intrasystem communications network included in the miscellaneous general hardware group and the on-line reading devices and automatic typewriters, none of which will have postdevelopmental software costs.

Costs attributed to the special reference graphic storage are not actually software costs, but rather the costs of installing the system and converting the relevant files to microfilm.

## *Rental vs. Purchase*

Since computer systems are frequently rented rather than purchased outright, appropriate estimates of total rental costs were developed for comparative purposes. The monthly rental of most computer equipment varies between $\frac{1}{40}$ and $\frac{1}{60}$ of the purchase price, based on a one-shift operation (176 hours) per month. A factor of $\frac{1}{45}$ of the purchase price was used in these estimates wherever specific rental data were not available. The total rental cost for the automated system estimated on this basis is approximately $372,000 per month or $4.36 million per year.

A premium is charged for additional operation above one shift per day; it was assumed, however, that all parts of the Library system would operate at least 1½ shifts, and the rental estimate was accordingly adjusted upward by a factor of 40 percent. This factor is derived from the present practices of the General Services Administration of the Federal Government. Applying this factor to the basic rental yields an estimated monthly rental of $522,000 and an annual rental of $6.25 million.

On this basis the estimated total purchase cost of the hardware ($26.3 million) is absorbed in about 4 years of rental. An amortization period of 4 years, however, is a slight understatement. Included in rental costs are service contract costs which, if paid for separately, are 2 percent of the purchase price per year. Even allowing for these included service benefits, the rental of the hardware equipment totals the purchase price in less than 5 years.

The rental of computer hardware is often justified on the basis of the rapid improvement in the state-of-the-art. Most of the hardware used in this system will be especially modified and designed to meet the Library's needs, and the 4-year period will barely provide opportunity for the automated Library to become fully operational. For these reasons, purchase rather than rental is assumed in the development of cost estimates in this report.

EXHIBIT 31.—SUMMARY OF BASIC HARDWARE AND SOFTWARE COSTS

| Hardware group | Description | Cost | | |
|---|---|---|---|---|
| | | Hardware [1] | Software [2] | Total |
| 100 | Internal processing group | $744,000 | $671,000 | $1,415,000 |
| 200 | Circulation and retrieval group | 969,000 | 786,000 | 1,755,000 |
| 300 | Central Catalog group | 5,000,000 | [3] 5,000,000 | 10,000,000 |
| 400 | Consoles | 5,440,000 | ([4]) | 5,440,000 |
| 500 | Editing processor | 3,363,000 | [3] 3,000,000 | 6,363,000 |
| 600 | Photocomposer | 500,000 | ([5]) | 500,000 |
| 700 | Business data processing and scheduling | 120,000 | 120,000 | 240,000 |
| 800 | Special reference graphic storage | 339,000 | [6] 25,000 | 364,000 |
| 900 | Miscellaneous general hardware | 208,000 | not applicable | 208,000 |
| | Total | 16,683,000 | 9,602,000 | 26,285,000 |

[1] Rounded to nearest thousand.

[2] Software costs, where applicable, were estimated to be equal to the programmable or logic-containing hardware, including storage files but excluding additional tape transport units and on-line reading devices or typewriters.

[3] Software costs for the automated search logic and the editing processor are not actually separable since the editing processor will control all input and some specialized outputs of the central catalog.

[4] Cost of programming set subroutines is included in purchase price of console hardware.

[5] This function will be controlled by the editing processor and all relevant software costs are attributed to the editing processor.

[6] Initial cost of committing clipping file to microfilm and organizing it was adapted from estimates provided by FMA, Inc.

# *Operating Costs of the Automated System*

## *Summary*

Annual salary costs of the automated system in 1972 total $4.2 million, of which $3.7 million is for salaries of the general Library staff and $0.5 million for equipment operators. Those areas of the Library whose operations are largely unaffected by automation are not included in this analysis, as explained in the text and exhibits which follow. The recurring annual costs of operating the basic hardware for the system is estimated at $317,600 for maintenance, consumable materials, and floor space, as detailed in Exhibit 34.

## *Personnel Costs*

Expenditures for the purchase of hardware and the development of software for the automated system are costs that must be met before the system can become operational. All costs incurred subsequent to this are defined as operating costs. In this report, a distinction is made between recurring costs directly relevant to basic hardware items and salary costs of Library personnel. The recurrent costs of basic hardware items include the costs of space, contract service maintenance, and consumable materials.

Recurrent hardware costs are fairly easy to analyze, but salary costs present certain problems. Since it is not the intention of this report to redesign the administrative organization of the Library, the present administrative structure has been used, insofar as is practical, to analyze the operating costs of the automated system. As mentioned previously, many divisions of the Library will be affected only to a minor degree by automation, since the processes automated comprise only a small portion of the total workload. Because of the difficulties inherent in abstracting minor work activities from a total work context, only the basic processing and circulation activities that form the core of the automated system are considered in detail. The relationship of automation to the remaining divisions is discussed, however, and some examples of savings that might be expected are pointed out.

Salary costs associated with the consoles, the special reference graphic storage equipment, and business data processing and scheduling are not presented in the analysis. The reason for the exclusion of consoles and special reference graphic storage equipment is that the automated equipment will affect only a small proportion of the total operations in which they are used. Business data processing and scheduling equipment is expected to produce substantial salary savings, but since this equipment will be installed in the Library well before the automated system as a whole is under way, an exhaustive analysis of salary costs in this area was not deemed necessary.

With the exception of cataloging, the use of the catalog, while crucial to many operations throughout the Library, does not occupy a major portion of the time spent in the total operation. In some cases consoles are assigned to divisions or areas in the Library not because they represent a significant time saving, but because they will be the only means of communication with the catalog and other files once these are no longer kept in a form which can be consulted manually.

One example of the kinds of uses to which the consoles are put, which cannot be readily analyzed in terms of salary costs, is in the General Reference and Bibliography Division of the Library. Reference librarians, while they make frequent use of the catalog, also spend a large part of their time preparing letters, consulting with readers, and answering the telephone requests. The use of the console, and the greater precision of the catalog, would afford substantial savings, but these savings would almost undoubtedly be reflected in more comprehensive answers to reference questions and in detailed treatment of a larger number of such questions, rather than a net reduction in personnel costs.

The special reference graphic storage equipment, while it would be of considerable assistance in the

Legislative Reference Service, would affect directly only the small percentage of the total employees who are responsible for maintaining the present clipping file. Indirectly it would have wider implications, since this file is the basis for many of the prepared reports and since almost all of the LRS research staff make some use of this file. The salary costs attributable to operating the new equipment roughly equals the salary costs presently associated with maintaining the clipping file. From the proportion of time spent, by the rest of the LRS staff using this file, it is possible to estimate approximately the total salary saving that may result from the installation of this new equipment. The estimated saving is approximately $40,000 per year, based on a reduction in average use time from 15 minutes to 5 minutes.

The three most important hardware groups in the automated system are those associated with internal processing, the central catalog, and circulation and retrieval. The importance of these hardware configurations, in terms of both complexity and total cost, accurately reflects those areas of the Library in which the automated system will make its most significant contribution. The conversion of the card catalog to digital form will have a pervasive effect on virtually every operating branch of the Library; however, with the exception of the three areas cited above, these effects will be for the most part indirect and may, in fact, be reflected more in improved and increased service rather than in basic changes in operating procedure or manpower requirements.

Exhibit 32 presents a summary of the operating costs of these three areas of operation. The companion exhibit (33) describes the hardware-oriented operating groups that are related both to the basic hardware groups discussed in Section VI and the present administrative divisions of the Library discussed in Section II.

The personnel costs of these groups are displayed both for the personnel required for equipment operation and for personnel required for the bibliographical and general library operations. The manpower requirements and salary costs associated with equipment operation were based on industry experience with comparable equipment configurations. Personnel included under this heading are programmers, maintenance technicians, operators of general-purpose computers, keypunch operators, and those console operators who are exclusively concerned with the console operation and are not called upon to exercise bibliographic judgment or to possess library skills.

Direct salary costs for programmers, computer operators, and maintenance technicians are based on rates presently prevailing in private industry and may represent a slight overstatement to the extent that Government salaries for these skills tend to be somewhat lower. Console operators and keypunch operator salaries are based on classifications (existing or comparable) presently prevailing within the Library. The direct salary costs associated with these positions are as follows:

| *Position* | *Annual salary* |
|---|---|
| Programming supervisor | $10,000 |
| Programmers | 8,000 |
| Computer technicians | 7,200 |
| Computer operators | 6,500 |
| Console operators | 5,255 |
| Keypunch operators | 4,345 |

EXHIBIT 32.—ANNUAL SALARY COSTS OF THE AUTOMATED SYSTEM: 1972

| Hardware-oriented operating groups [1] | Number of man-years | | | Salary cost [3] | | |
|---|---|---|---|---|---|---|
| | Equipment operation [2] | General library staff | Total | Equipment operation [2] | General library Staff | Total (dollars) |
| Cataloging and catalog maintenance | [4] 47 | 256 | 303 | 318,100 | 1,845,300 | 2,163,400 |
| Acquisition and internal processing | 8 | 170 | 178 | 48,700 | 964,400 | 1,013,100 |
| Circulation, retrieval, and storage | 14 | 192 | 206 | 85,200 | 907,200 | 992,400 |
| Total | 69 | 618 | 687 | 452,000 | 3,716,900 | 4,168,900 |

[1] For explanation and description of hardware-oriented group see Exhibit 33.

[2] Includes technicians, operators of general-purpose computers, keypunch operators and console operators who operate consoles full time and are not required to exercise bibliographic judgment.

[3] Includes 7½ percent fringe benefit factor.

[4] Includes all programmers for the system.

The estimation of the remaining labor costs involved a more complex procedure. Using work units identical with or adapted from a study prepared by the Planning Research Corporation for the Library of Congress in December 1961, output per man-year was estimated in each of the relevant activities. The resultant outputs per man-year were adjusted upward to take into account the increased productivity expected to result from automation and were then applied to the volume projected for 1972 in Section V. From this an estimate was obtained for the number of man-years required for each operation in the automated system. In all cases the number of man-years for a given operation was rounded upward to the nearest whole number.

The factors used for adjusting the present output per man-year to the output per man-year under automation were the reciprocals of work reduction coefficients. The primary work reduction coefficient was 60 percent, which reflected the proportion of time saved in general searching in the main card catalogs. This figure was derived from a recent survey of searching activity conducted by the Office of the Information Systems Specialist. In this study it was found that almost 60 percent of the total time in searching new titles was spent walking from one file to another in following cross-references or checking alternative entries. This walk time would be eliminated by the use of the console; however, the elimination of walk time is only one part of the savings in catalog searching, since the automated catalog should be more refined, complete, and less susceptible to error than the present card catalogs.

The 60 percent work reduction coefficient was applied only to complete catalog searching operations; smaller work reduction coefficients were attributed to more rationalized searching activities where only a single file was used and walk time or lookup time was reduced. For instance, the search time associated with serial record posting was only assumed to have a work reduction coefficient of 30 percent.

Once the number of man-years required for the operations in the three major areas was determined, the salary costs were estimated using the present salary structure. No allowance was made in this case, nor in the case of personnel involved with the equipment operation, for salary changes. In addition, all cost estimates are in 1962 dollars.

The salary costs estimated in Exhibit 32 include a 7½ percent fringe benefit factor, since this factor was used in the previous study of Library operations conducted by the Planning Research Corporation.

## *Recurrent Hardware-Related Costs*

Three basic kinds of recurring costs are analyzed in this report: costs of providing a suitable environment (space costs), service contract costs, and consumable materials costs. The day-to-day maintenance cost associated with the equipment is subsumed under the general labor costs of operating the equipment, but the costs for service provided by the manufacturer are treated separately, since these are usually handled on a contract basis. Power costs are minimal in the total operation and were therefore excluded. For example, the electricity to run a large general-purpose

Exhibit 33.—Descripton of Hardware-Oriented Groups

| Hardware-oriented operating groups [1] | Comparable units in present administrative structure [2] | Relevant basic hardware group |
|---|---|---|
| Cataloging and Catalog Maintenance. | Descriptive Cataloging Division.<br>Subject Cataloging Division.<br>Decimal Classification Office.<br>Catalog Maintenance Division.<br>Union Catalog Division.<br>New Serial Titles Section. | Central Catalog Group.<br>Consoles (in part). |
| Acquisition and Processing. | Order Division.<br>Serial Record Division.[3]<br>Exchange and Gift. | Processing Group. |
| Circulation, Retrieval, and Storage. | Loan Division.<br>Stack and Reader Division.<br>Serial Division. | Circulation and Retrieval Group. |

[1] As used in Exhibit 32.
[2] See Section III for description and position.
[3] Excluding the processing and publication of *New Serial Titles* in the New Serial Titles Section.

computer with a fairly complete line of support equipment costs only about 30 cents per hour. Even allowing for several such computers with 4,000 hours of operation per year, the total power cost would still amount to a few thousands of dollars.

Exhibit 34 is a summary of these recurring costs tabulated by the basic hardware groups designated in Section VI. The first column indicates the amount and the cost of the floorspace required for various operations. While it is probable that the installation of the automated system would result in a net savings of floorspace over all, the cost of floorspace specifically required for the installation of the equipment is tabulated for two reasons. The first reason is that, while all floorspace used for any operation has some cost, the cost of providing floorspace for expensive computer equipment is higher, since temperature and humidity must be controlled. A second reason is that the space savings permitted by the installation of the automated equipment will not be immediately realized, since existing card catalogs, etc., will still have to be maintained for an indefinite period of time. It was felt that the inclusion of floorspace costs would provide a more generous estimate of recurrent costs.

For the most part, service contract costs are estimated from computer manufacturer pricelists and a representative sample of service costs for the proposed equipment was tabulated. These contract costs vary increasingly with the time elapsed since installation of the equipment, with increments at 3-year intervals; therefore, the average of the entire 9-year period provided in the available pricelists was used. This average rate (determined from the sample to be 1.6 percent of the purchase price per year) works out to be somewhat greater than the rate for the middle period. This rate has been applied to the purchase prices developed in Section VI for all the hardware groups except the console and the graphic storage equipment.

EXHIBIT 34.—RECURRING ANNUAL COSTS OF BASIC HARDWARE ITEMS

| Basic hardware group [1] | Floorspace | | Service contract cost [4] | Consumable materials cost | Total recurring cost |
|---|---|---|---|---|---|
| | Square feet required [2] | Total cost cost [3] | | | |
| 100 Internal processing group | 1,000 | $5,000 | $11,900 | [5] $500 | $17,400 |
| 200 Circulation and retrieval group | 1,000 | 5,000 | 15,500 | [5] 500 | 21,000 |
| 300 Central catalog group [6] | 2,500 | 12,500 | 80,000 | ..... | 92,500 |
| 400 Consoles [7] | ..... | ..... | [8] 544,800 | ..... | 54,400 |
| 500 Editing processor | 2,500 | 12,500 | 67,300 | [9] 3,000 | 82,800 |
| 600 Photocomposer | 700 | 3,500 | 15,000 | [10] 8,700 | 27,200 |
| 700 Business data processing and scheduling | 500 | 2,500 | 1,900 | [5] 500 | 4,900 |
| 800 Graphic storage equipment | 400 | 2,000 | [11] 9,000 | [12] 3,100 | 14,100 |
| 900 Miscellaneous general hardware | ..... | ..... | 3,300 | ..... | 3,300 |
| Total | ..... | 43,000 | 258,300 | 16,300 | 317,600 |

[1] Numbers and descriptions keyed to exhibits in Section VI.

[2] Estimated size of room required to house equipment, desks, and tables required by operators and users.

[3] Estimated at $5.00 per square foot including air conditioning and other space-related costs.

[4] Service contract costs for computers and related equipment range from 1 percent to 4 percent of purchase price, depending in part on the extent to which high-speed mechanical devices are used. A representative sample of types of equipment used in the automated system indicated a mean service contract rate of 1.6 percent of purchase price, which is the estimate used here unless otherwise noted.

[5] Based on an annual replacement rate of ten 2,400-foot reels of magnetic tape. Punched cards and paper costs are ignored as negligible.

[6] All consumable materials including tapes for programs relating to the catalog are subsumed under the editing processor.

[7] No additional floorspace is allocated to consoles, since it is assumed that files and desks replaced will more than compensate for the console space requirements.

[8] Service contract cost is estimated at a lower rate of 1.0 percent, since it is assumed that redundancy (more than 160 consoles throughout the Library) will provide substantial economies of scale.

[9] Based on annual replacement of fifty 2,400-foot reels of magnetic tape at $50 plus $200 in paper and punched card costs.

[10] Based on an estimated materials cost of 10 cents per page generated. Assuming an average of 600 words per page, an annual output of 50 million words would generate approximately 87,000 pages.

[11] Based on service contract rates charged by FMA, Inc., for Filesearch equipment.

[12] Costs of film, punched cards, and film reels estimated from data provided by FMA, Inc.

SECTION VIII

# *Costs of Conversion to Digital-Form Storage*

## *Summary*

Several alternatives are possible with respect to the files converted to digital form. Conversion of minimum basic files (53 million words) would cost $1.2 million; basic files plus the Library of Congress catalog (426 million words) would cost $2.4 million; basic files plus the National Union Catalog, which includes the LC catalog (699 million words), would cost $3.3 million. The cost does not vary directly with the conversion load since the conversion method considered (stenotypy) has a fixed cost of $1 million. Stenotypy is considered practical because the magnitude of the conversion load permits development of automated devices, such as an automatic conversion dictionary, for the conversion operation. The conversion cost, including 100 percent verification in proofreading, is estimated at ⅓ cent per word.

## *The Conversion Process*

The establishment of an automated system for the Library would require the complete or partial conversion of card catalogs and other files and records to machine-readable (digital) form. For many of the automated functions, the relevant records, such as the reference files associated with acquisition, processing, and circulation, would require conversion to digital form even before the system can operate. Conversion of all or part of the contents of the card catalogs to machine-readable form is not, however, a prerequisite for operating the system. When the system begins operation, the automated catalog can be empty, partially complete, or complete, although the first two alternatives would involve the inconvenience and cost of using two catalogs concurrently. In any case, it is probable that, even if the conversion of the card catalog to digital form were not complete before the system began operation, provision would be made for complete conversion within a reasonable length of time.

For purposes of estimating requirements and costs, it is assumed throughout this report that all catalog cards and other operating files have been converted to digital form and stored in the appropriate automated files. This assumption is a necessary one if the resulting cost estimates are to apply to a completed automated system and not to some transitional system which will require further modification or extension when catalog conversion is completed.

In this section, general cost estimates, based on present technology in this area, are developed both for the conversion of records required for system operation and for complete conversion of the Main Catalog. It was assumed that the conversion process would not be accomplished automatically through the use of electronic equipment capable of translating printed material to machine-readable form without any human intermediary, since at the present time automatic conversion devices are capable of dealing with only a single typeface and therefore would not be suitable.

There are at the present time few man-machine processes for converting textual information to machine-readable form, other than typewriting (keypunching) operations. One process, recently developed specifically for rapid conversion of a large volume of varied information, involves the use of stenotype transcription and automated lexical processing. The cost estimates developed here are based on this method of conversion; however, this stenotype conversion is used only as a reasonable basis for estimating approximate costs, and a much more detailed analysis is needed before a determination can be made of the specific method to be used.

Both fixed and variable costs are associated with stenotype conversion. The fixed costs apply to the automated dictionary through which stenotype copy is processed and to the related electronic equipment.

This automated dictionary is analogous to the dictionaries used in automatic translation of foreign languages. In this method the stenotyped copy is compared in a large dictionary, words are spelled out, arranged and edited, and transferred in digital form to magnetic tape or other digital storage.

While the particular steps in this conversion process may not be equally applicable to all facets of the conversion problem, the estimates presented below are considered representative of feasible alternative conversion methods. The costs associated with the automated dictionary, its support equipment, and operation are esesntially fixed and do not vary appreciably with the volume of information processed or the specific input process utilized. The variable costs of stenotype conversion are associated with the stenotype transcription and with proofreading the machine output. Salary cost for the transcription and proofreading are largely proportional to the amount of data converted, although there are certain limited fixed initial training costs.

The fixed costs associated with the automated dictionary and support equipment are approximately $1 million. It should be noted that this amount includes an automated dictionary for the English language only, since at the present time automated lexical processing is practical only for the English and Russian languages, these being the languages for which automated dictionaries have been compiled. Accurate estimates of the cost of generating other dictionaries are not available. Furthermore, not even approximate estimates of the composition of the Main Catalog and other files, by language, are available. The relative importance of foreign languages in the various files and catalogs is a problem of special importance which will require considerable study. In view of the lack of pertinent information, and for purposes of simplifying the analysis, the entire conversion process was assumed to require only the English language.

The basic unit in the estimation of variable costs is the word, rather than the character or the bit, since both the stenotype transcription and the proofreading of machine-generated copy are specifically concerned with word units. The conversion cost, including both stenotyping and 100 percent verification in proofreading, is ⅓ cent per word.[1]

[1] Projected estimate given the Planning Research Corp. by a group active in this field.

## *Cost of File Conversion*

Exhibit 35 lists five files that must be converted to digital form before the relevant automated functions can become operational. The first three files (monograph, serial, and exchange) provide basic working information on sources for new acquisitions and for serial titles to which receipts must be posted. The initial inputs to the not-on-shelf file will be transfers of charge records for items in use in the Library (the present Central Charge File). The complete conversion of the catalog authority file is dictated by the importance of the file as an aid to cataloging and by the fact that the established author and subject-heading entries are basic to the organization of information in the automated catalogs. The number of words in the National Union Catalog is also shown (the Library of Congress Main Catalog is given for purposes of comparison, although each main entry in the Main Catalog is included in the National Union Catalog).

Exhibit 36 shows the costs of conversion for the various files and alternative combinations of files. The lowest cost estimate is $1,177,000 for conversion of only those files necessary for operation of the system. The highest estimate is $3,331,000 for conversion of the entire contents of the National Union Catalog plus the required files. Because of the relatively large fixed-cost elements, the proportional increase in total cost is much smaller than the proportional increase in the volume of information processed. Hence, the processing of over 2.3 million words costs only three times as much as the processing of 177,000 words, a volume only $\frac{1}{13}$ as large.

The cost of conversion should not be regarded simply as the cost of transferring identical records from one format to another, since the conversion of records to digital form can provide many important byproducts at little or no additional cost. All data to be converted must pass through the automated conversion dictionary and the computer associated with it, a process that would permit many kinds of lists and tabulations to be generated automatically as part of the conversion process. In the conversion of catalog card information, for example, key title words could be tabulated and stored along with the appropriate item identification numbers. Thus a title-word index could be produced for all prior holdings at only a

small additional cost for off-line equipment and tape.

Similarly, the conversion of the catalog authority file would provide automatic editing and cross-checking of author and subject headings for catalog card conversion. Names or terms particularly subject to misinterpretation or misfiling could be specifically noted and automatically cross-checked in any of the files. Another important byproduct of this process might be the flagging of duplicate, incomplete, or ambiguous entries for special examination.

EXHIBIT 35.—FILES TO BE CONVERTED TO DIGITAL FORM

| Description | Number of characters | Number of words [1] |
|---|---|---|
| A. *Required files:* [2] | | |
| Standing order and vendor files in the monograph processing file (see Exhibit 17) | 3, 600, 000 | 720, 000 |
| Initial size of serial record file (see Exhibit 18) | 66, 331, 000 | 13, 266, 000 |
| Complete list of exchange and gift sources in the exchange record (see Exhibit 19) | 13, 800, 000 | 2, 760, 000 |
| Transfer of Central Charge File records to not-on-shelf file (see Exhibit 21) | 3, 200, 000 | 640, 000 |
| Catalog authority file (see Exhibit 23) | 178, 765, 000 | 35, 753, 000 |
| Total | 265, 696, 000 | 53, 139, 000 |
| B. *Prior holdings:* | | |
| National Union Catalog: | | |
| 11,750,000 titles [3] at 275 characters [4] | 3, 321, 250, 000 | 646, 250, 000 |
| Library of Congress Main Catalog: [5] | | |
| 6,600,000 titles [6] at 275 characters [4] | 1, 865, 000, 000 | 373, 000, 000 |

[1] Estimated as 1/5 the number of characters and rounded to the nearest thousand.

[2] Files that must be converted to digital form before the system can become fully operational.

[3] Based on estimates of present size and projected increase given in Exhibit 14.

[4] See note (3) of Exhibit 14.

[5] The contents of this catalog are included in the N.U.C.

[6] Based on an estimated maximum of 5.5 million titles in 1961 plus the increase projected by 1972; see note (2) of Exhibit 14.

EXHIBIT 36.—COSTS OF CONVERSION OF EXISTING FILES TO DIGITAL FORM

| Basic files | Number of words | Fixed cost [1] | Variable cost [2] | Total cost |
|---|---|---|---|---|
| 1. Required files (listed under A in Exhibit 35) | 53, 139, 000 | $1, 000, 000 | $177, 000 | $1, 177, 000 |
| 2. National Union Catalog | 646, 250, 000 | 1, 000, 000 | 2, 154, 000 | 3, 154, 000 |
| 3. Library of Congress Main Catalog only | 373, 000, 000 | 1, 000, 000 | 1, 242, 000 | 2, 242, 000 |
| *Combined files* | | | | |
| 4. Required files plus National Union Catalog (variable cost of each plus fixed cost) | 699, 389, 000 | 1, 000, 000 | 2, 331, 000 | 3, 331, 000 |
| 5. Required files plus LC Main Catalog only (variable cost of each plus fixed cost) | 426, 139, 000 | 1, 000, 000 | 1, 420, 000 | 2, 420, 000 |

[1] Cost of automated dictionary and support equipment.

[2] Estimated at 1/3 cent per word (see text for included costs).

SECTION IX

# *Cost Projection for the Manual System*

## *Summary*

Salary costs for the computer system are compared to salary costs for maintaining the present manual operations for the increased workloads projected for 1972. Results indicate that salary costs will be higher for the manual system—$5 million compared to $4.2 million. This may be a particularly significant finding, since libraries are having to compete more and more with other agencies for the highly trained manpower they require. In addition to matching the manual system operationally, the automated system provides many services not possible in a manual operation. No attempt was made to estimate what it would cost the Library to perform these same services manually; hence the figures above are understated for the manual operation.

## *Introduction*

Salary cost estimates for certain divisions and operations in the Library were developed for the automated system in Section VII. In order to provide a basis for comparison, similar salary cost estimates are developed in this section for the present manual system for the same divisions and operations. Costs are presented for the same hardware-oriented operating groups used in Exhibit 32 and described in Exhibit 33.

The unit costs estimates used in this analysis are taken from a previous study of the Library's operations conducted by the Planning Research Corporation. This study, entitled *Cost Estimates of Selected Information Storage and Retrieval Activities,* was completed in December 1961 after several weeks of investigation in the Library. Specific work units or work unit equivalents were developed to measure the amount and type of activities performed in the Library. These work units were defined for the various administrative divisions of the Library, and all salaries in each division studied were allocated to at least one of the established work units.

Although the cost estimates are developed for administrative groups identical to those used for the automated system, the two sets of estimates are only roughly comparable in that no attempt has been made to separate those additional services and refinements that the automated system will provide.

## *Salary Cost Estimates*

The salary cost estimates shown in Exhibits 37 through 39 are based on the workload projections made for the automated system. In some cases the various statistics and assumptions used to define the automated system are not as well suited for analysis of the nonautomated system, with the result that the salary costs derived are not necessarily identical to the estimates that might have been developed had the primary object of this report been to project costs for a nonautomated system. For example, in some instances, particularly for those operations eliminated by automation, projections were not developed. In such cases the estimates used in Exhibit 38 were adapted from available projections, usually by applying ratios observed for 1961 operations. All such extrapolations and adjustments have been noted in the exhibits. Salary costs in this case, as in the costs for the automated system, include a fringe-benefit factor of 7½ percent, since this factor was included in the original unit costs developed in the December 1961 study.

Salary costs associated with the catalog and catalog maintenance hardware-oriented group are tabulated in Exhibit 38. The total salary cost for this operating group in 1972 is estimated to be $2.4 million. Comparable tabulations of salary costs for the acquisition and internal processing group and for the circulation, retrieval, and storage group are presented in Exhibits

37 and 39, respectively. Much of the routine work of the Order Division will become completely automatic and the operations of the Exchange and Gift Division may be drastically revised by virtue of electronic access to all acquisition records. Consequently, there are in the automated system no separately identifiable equivalents of the present operations of these two divisions. In view of this, the salary costs in Exhibit 37 for these two divisions have been estimated somewhat arbitrarily at twice the 1961 level. This estimate parallels the estimated 100 percent increase in transactions volume used for purposes of establishing the requirements of the exchange record in Exhibit 19. All cases in which the unit cost or estimated volume reflect adjustments or extrapolations from estimates used elsewhere in this report are noted.

The resulting salary cost estimates for the three hardware-oriented groups are as follows:

| Operating group | 1972 salary costs manual operation | 1972 salary costs automated system |
|---|---|---|
| Cataloging and Catalog Maintenance | $2, 375, 000 | $2, 163, 400 |
| Acquisition and Internal Processing | 1, 297, 400 | 1, 013, 100 |
| Circulation, Retrieval, and Storage | 1, 313, 310 | 992, 000 |
| | 4, 985, 710 | 4, 168, 500 |

For a discussion of the benefits resulting from automation, for which a comparable manual system was not projected or costed out, see Section I, pages 33 to 34.

EXHIBIT 37.—SALARY COSTS OF THE MANUAL SYSTEM IN 1972: ACQUISITION AND INTERNAL PROCESSING

| Administrative unit | Equivalent work units | Salary cost per unit [1] | Estimated volume | Total salary cost [1] |
|---|---|---|---|---|
| Order Division | All activities assumed double 1961 [2] | ..... | ..... | $486, 192 |
| Serial Record Division | Pieces searched [3] | $0. 087 | [4] 2, 100, 000 | 182, 700 |
| | New entries prepared [3] | 3. 090 | [4] 20, 000 | 61, 800 |
| | Other operations [5] | ..... | presumed constant | 45, 004 |
| Exchange and Gift Division | All activities assumed double 1961 [6] | ..... | ..... | 521, 720 |
| Total salary cost | | | | 1, 297, 416 |

[1] Includes 7½ percent fringe benefit factor.
[2] Since this operation will be included in the larger monograph acquisitions and serial acquisitions functions, the work units defined in the previous study were not too relevant to the automated system. In allocating operating costs in the automated system, a general increase of 100 percent (double) by 1972 was assumed. The same assumption is used here for comparability.
[3] Work units defined in earlier cost study.
[4] Estimates confirmed by division personnel.
[5] Includes physical maintenance of files, physical transfer of materials, etc.
[6] This division will also be changed radically by automation, making individual work units previously defined irrelevant. In determining operating costs for the automated system, provision was made for twice the present level of operations. The same assumption is made here.

EXHIBIT 38.—SALARY COSTS OF THE MANUAL SYSTEM IN 1972: CATALOGING AND CATALOG MAINTENANCE

| Administrative units | Equivalent work units [1] | Salary cost per unit [2] | Estimated volume [3] | Total salary cost [4] |
|---|---|---|---|---|
| Descriptive Cataloging Division | New titles | [5] $8.52 | 120,000 | $1,022,760 |
| | Titles recataloged or edited | [6] 3.29 | 31,000 | 102,021 |
| Subject Cataloging Division [7] | Titles assigned subject headings and classification. | [8] 2.91 | [9] 102,000 | 297,228 |
| Decimal Classification Office | Titles classified | .78 | 40,000 | 31,200 |
| Other Operations [10] | | | presumed constant | 46,225 |
| Catalog Maintenance Division | Cards prepared for filing | .04 | [11] 2,446,000 | 97,840 |
| | Cards filed | .06 | [11] 1,892,000 | 113,520 |
| Other Operations [12] | | | presumed constant | 150,768 |
| Union Catalog Division: | | | | |
| Current File | Cards received | .19 | [13] 1,368,000 | 259,920 |
| Past File | Cards received | .11 | [13] 1,032,000 | 113,520 |
| Information Service | | | presumed constant | 51,388 |
| New Serial Titles Section [14] | Titles searched | .12 | [15] 158,000 | 18,960 |
| | Titles edited | 3.17 | [16] 22,000 | 69,740 |
| Total cost | | | | 2,375,090 |

[1] The work units are equivalent to thse developed in previous cost study by Planning Research Corporation for the Library of Congress. In some cases these work units have been adapted to simplify the exposition, such adaptationsare noted in the footnotes.

[2] Salary costs per unit are taken directly from the earlier cost study, unless otherwise noted.

[3] Unless otherwise noted, the estimates of volume for the various operations are taken from Exhibit 15 to insure comparability with the cost analysis of the automated system.

[4] Total salary costs and salary costs per unit include a 7½ percent fringe benefit factor in previous cost study.

[5] This estimate is sum of 3 unit costs for component procedures in descriptive cataloging: searching ($0.86), preliminary cataloging ($1.08), and descriptive cataloging ($6.58).

[6] Only descriptive cataloging unit costs apply in this case, since searching and preliminary cataloging are not required for cooperative editing or recataloging. In the earlier cost study it was estimated that editing and recataloging required only ½ the direct labor of full cataloging. For this reason the unit cost estimate used here is equal to ½ the descriptive cataloging unit cost cited in note (5).

[7] The shelflisting and labeling operations carried on within the subject cataloging division are, for purposes of this analysis, assigned to the circulation, retrieval, and storage group because of the inclusion of the location directory in that group.

[8] Unit cost refers only to subject analysis and classification for each title, and does not include shelflisting and labeling.

[9] A study of the Main Catalog conducted by the Office of the Information Systems Specialist indicates that approximately 85 percent of the titles in that catalog had subject headings. Hence, only 85 percent of the 120,000 new titles (102,000) are assumed to require subject cataloging.

[10] Other operations in this case refer to the editing and preparation of the published *Dewey Decimal Classification* and other manuals, etc. These activities do not appear to be directly related to the number of titles classified.

[11] Statistics on cards filed were not developed for analysis of the automated system since actual cards will not be required. For purposes of this cost analysis, however, an increase of ⅓ over 1961 levels was assumed, based on the percentage increase projected for new titles cataloged.

[12] Other operations listed in this division include preparation of the published NUC catalog and answering reference queries.

[13] The total number of NUC reports projected for 1972 is 2.4 million (see Exhibit 15). In 1961, reports processed in the current file (post-1955) account for 57 percent of the total reports. Since the cutoff date for the current file in general progresses each year, this ratio is assumed to remain constant, and the total volume of reports projected for 1972 is divided accordingly between the current and past files.

[14] *New Serial Titles* is a publication of the Serial Record Division, but it has been included in this operating group for purposes of this analysis, since the operations required for its preparation are analogous to those involved in the Union Catalog Division.

[15] See Exhibit 18.

[16] Estimate confirmed by Serial Record Division personnel.

EXHIBIT 39.—SALARY COSTS OF THE MANUAL SYSTEM IN 1972: CIRCULATION, RETRIEVAL, AND STORAGE

| Administrative unit | Estimated salary cost, 1972 |
|---|---|
| Loan Division | [1] $272,224 |
| Stack and Reader Division | [1] 396,561 |
| Serial Division | 275,645 |
| Subject Cataloging Division: | |
| Shelflisting | [2] 282,960 |
| Labeling | [3] 85,920 |
| Total salary cost | 1,313,310 |

[1] Salary costs for these divisions are estimated at FY 1961 levels. In developing costs for the automated system, no projections were made either because the volume was determined by exogenous factors or because historical statistics indicated no significant trend (see Exhibit 13).

[2] For purposes of this estimate, the estimated number of new titles in 1972 (120,000, see Exhibit 15) was assumed to be equal to the number of titles shelflisted. The earlier cost study indicated a cost of $2.36 per title shelflisted (120,000x $2.36=$282,960).

[3] The ratio of volumes labeled to titles shelflisted was assumed to be the same in 1972 as 1961. This ratio of approximately 4:1 was applied to 120,000 titles shelflisted estimated above for 1972: cost per volume labeled, according to the earlier cost study, is $0.18 (480,000x$0.18=$85,920).

U.S. GOVERNMENT PRINTING OFFICE:1964

LIBRARY OF CONGRESS
0 020 567 494 9